

日 本 国 特 許 庁
JAPAN PATENT OFFICE

12.11.03

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日
Date of Application: 2003年 1月21日

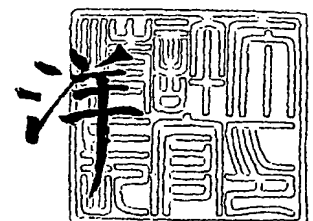
出 願 番 号
Application Number: 特願2003-012545
[ST. 10/C]: [JP2003-012545]

出 願 人
Applicant(s): 西山 修平
株式会社クエスト

2005年 1月13日

特許庁長官
Commissioner,
Japan Patent Office

小 川



【書類名】 特許願

【整理番号】 NISIYM-001

【あて先】 特許庁長官 殿

【国際特許分類】 G06F 13/00
G06F 15/16
G06F 17/30

【発明者】

【住所又は居所】 千葉県浦安市明海 1 4 番地海園の街 2 - 8 0 5 号

【氏名】 西山 修平

【特許出願人】

【住所又は居所】 千葉県浦安市明海 1 4 番地海園の街 2 - 8 0 5 号

【氏名又は名称】 西山 修平

【代理人】

【識別番号】 100109553

【弁理士】

【氏名又は名称】 工藤 一郎

【手数料の表示】

【予納台帳番号】 100322

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 分散型データベースシステム

【特許請求の範囲】

【請求項 1】 ネットワーク上に配置されたデータベースを管理するデータベース管理装置と、前記データベース管理装置のデータベースを分散するためのトポロジ管理サーバ装置と、前記トポロジ管理サーバ装置が前記ネットワーク上で管理する管理ドメイン内に配置された複数の計算機と、を備える分散型データベースシステムであって、

前記トポロジ管理サーバ装置は、

前記データベース管理装置により管理されるデータベースオブジェクトを識別するための情報であるデータベースオブジェクト識別子と、そのデータベースオブジェクトが管理されるデータベース管理装置を識別するためのデータベース管理装置識別子と、を対応付けた情報を含む情報であるトポロジ情報を格納するトポロジ情報格納部と、

前記データベースオブジェクト識別子で識別されるデータベースオブジェクトをキャッシュするために前記計算機の少なくとも一以上から発せられる前記データベースオブジェクト識別子を含むキャッシュ要求を受信するキャッシュ要求受信部と、

前記キャッシュ要求受信部で受信されたキャッシュ要求に含まれるデータベースオブジェクト識別子に基づいて前記トポロジ情報格納部から対応するデータベース管理装置識別子を取得するデータベース管理装置識別子取得部と、

前記データベース管理装置識別子取得部で取得されたデータベース管理装置識別子で識別されるデータベース管理装置に対して、前記キャッシュ要求を転送するキャッシュ要求転送部と、

を有し、

前記計算機は、

キャッシュ要求を送信するキャッシュ要求送信部と、

キャッシュ要求送信部でのキャッシュ要求の送信に応じて返信されるデータベースオブジェクトを受信するデータベースオブジェクト受信部と、

データベースオブジェクト受信部で受信されたデータベースオブジェクトをキャッシュするデータベースオブジェクトキャッシュ部と、
を有し、

前記データベース管理装置は、

トポロジ管理サーバ装置から転送されるキャッシュ要求を受信するキャッシュ要求受信部と、

キャッシュ要求受信部で受信されたキャッシュ要求に応じてデータベースオブジェクトを複製して送信するデータベースオブジェクト複製送信部と、
を有する、

分散型データベースシステム。

【請求項 2】 前記計算機は、データベースオブジェクトキャッシュ部にキャッシュされたデータベースオブジェクトを参照しながら計算を行なうためのプログラムを受信するプログラム受信部を有する請求項 1 に記載の分散型データベースシステム。

【請求項 3】 前記計算機は、自身の計算負荷を検出する計算負荷検出部を有し、

キャッシュ要求送信部は、計算負荷検出部で検出された計算負荷に基づいてキャッシュ要求を送信することを特徴とする請求項 1 または 2 に記載の分散型データベースシステム。

【請求項 4】 キャッシュ要求送信部は、計算機が起動するとキャッシュ要求を送信することを特徴とする請求項 1 または 2 に記載の分散型データベースシステム。

【請求項 5】 キャッシュ要求には、データベースオブジェクトキャッシュ部がキャッシュできるデータベースオブジェクトの量又は／及びデータ種別に関する情報が含まれることを特徴とする請求項 1 ないし 4 のいずれかーに記載の分散型データベースシステム。

【請求項 6】 キャッシュ要求には、データベースオブジェクトキャッシュ部がデータベースオブジェクトをキャッシュできる時間帯に関する情報が含まれることを特徴とする請求項 1 ないし 4 のいずれかーに記載の分散型データベースシ

テム。

【請求項 7】 キャッシュ要求には、計算機の有する計算能力に関する情報が含まれることを特徴とする請求項 1 ないし 4 のいずれかに記載の分散型データベースシステム。

【請求項 8】 キャッシュ要求には、計算機がデータベースオブジェクトをキャッシュすることに対する対価に関する情報が含まれる請求項 1 ないし 4 のいずれかに記載の分散型データベースシステム。

【請求項 9】 トポロジ情報は、さらに、データベースオブジェクト識別子に、データベースオブジェクトがキャッシュされた計算機の識別子を対応づけ、

前記トポロジ管理サーバ装置は、

データベースオブジェクトが計算機にキャッシュされたことを示す情報であるキャッシュ済情報を受信するキャッシュ済情報受信部と、

キャッシュ済情報受信部で受信されたキャッシュ済情報に基づいて、トポロジ情報格納部に格納されたトポロジ情報のキャッシュ済み情報を最新の状態に更新するトポロジ情報キャッシュ更新部と、

を有する請求項 1 に記載の分散型データベースシステム。

【請求項 10】 前記トポロジ情報は、さらにデータベースオブジェクト識別子に、データベースオブジェクトに操作されたロックに関する情報であるロック情報を対応付け、

前記トポロジ管理サーバ装置は、

ロック情報を受信するロック操作情報受信部と、

ロック操作情報受信部で受信されたロック情報に基づいて、トポロジ情報格納部に格納されたトポロジ情報のロック情報を最新の状態に更新するトポロジ情報ロック更新部と、

を有する請求項 9 に記載の分散型データベースシステム。

【請求項 11】 前記トポロジ管理サーバ装置は、複数あり、かつ、前記トポロジ管理サーバ装置は、ネットワークにより通信が可能な他の管理ドメインを管理する他トポロジ管理サーバ装置とトポロジ情報を交換するトポロジ情報交換部を有する請求項 9 または 10 に記載の分散型データベースシステム。

【請求項 12】前記データベース管理装置は、

自らが保持するデータベースオブジェクトに対して更新操作を行なうと、データベースオブジェクトをキャッシュしている計算機であるクライアント装置に向けて、キャッシュされているデータベースオブジェクトの更新操作の命令である更新操作命令を送信する更新操作命令送信部を有し、

前記計算機は、

更新操作命令を受信する変更操作命令受信部と、

更新操作命令受信部で受信された更新操作命令に基づいて、データベースオブジェクトキャッシュ部にキャッシュされたデータベースオブジェクトを更新するデータベースオブジェクト更新操作部を有する請求項 2、9、11 のいずれかに記載の分散型データベースシステム。

【請求項 13】前記データベース管理装置は、

データベースオブジェクトに対する更新操作に関する情報である更新操作情報を受信する更新操作情報受信部と、

更新操作情報受信部で受信された更新操作情報に基づいて自らが保持するデータベースオブジェクトに対して更新操作を行なう更新操作部と、

前記更新操作情報に基づいて、データベースオブジェクトをキャッシュしている計算機であるクライアント装置に向けて、キャッシュされているデータベースオブジェクトの更新操作の命令である更新操作命令を送信する更新操作命令送信部を有し、

前記計算機は、

更新操作情報を送信する更新操作情報送信部と、

更新操作命令を受信する更新操作命令受信部と、

更新操作命令受信部で受信された更新操作命令に基づいて、データベースオブジェクトキャッシュ部にキャッシュされたデータベースオブジェクトを更新するデータベースオブジェクト更新操作部を有する請求項 2、9、11 のいずれかに記載の分散型データベースシステム。

【請求項 14】前記データベース管理装置で管理されるデータベースオブジェクトに対する更新を表わすジャーナルを受信するジャーナル受信部と、

ジャーナル受信部で受信されたジャーナルを蓄積するジャーナル蓄積部と、
ジャーナル蓄積部で蓄積したジャーナルをリプレイするジャーナルリプレイ部
と、

前記ジャーナルリプレイ部でリプレイされたジャーナルに基づいて生成された
スナップショットを保持するスナップショット保持部と、

ドメインが障害を蒙ったとき前記スナップショットから障害ドメインのリカバ
リのための処理を行なう機能を有するリカバリ部と、

を有するジャーナル管理サーバ装置をさらに備え、

前記データベース管理装置は、

ジャーナルを送信するジャーナル送信部を有する請求項 1 に記載の分散型デー
タベースシステム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、分散データベースシステム及びそれを用いたグリッドコンピューテ
ィングシステムに関する。

【0002】

【従来の技術】

従来の商用化されたりレーショナルデータベースシステムにおけるデータの分
散は、二相コミットとレプリケーションとによって実現されてきた。また、デー
タベースの格納媒体はハードディスク装置であり、バックアップの採取はデー
タベースの停止を前提としていた。

【0003】

二相コミットとは、正規化されて参照／被参照関係にあるテーブルのセルの間
で、被参照テーブルにあるセルの値の変更やそのセルが属する行の削除を行な
おうとするとき、参照テーブルが複数のホストコンピュータに分散していること
を想定し、存在しない被参照セルを参照するセルが発生しないよう、一度自ホ
ストコンピュータ上の当該被参照テーブル上でチェックし、参照セルが無い場
合に、その更新を仮コミットし、その後、他のホストコンピュータ上の参照テ
ーブルを

チェックし、そこでも参照セルが無い場合に、はじめてその更新を本コミットする仕組みで、二段階に渡ってコミットするため二相コミットと言われる。

【0004】

二相コミットについては、システムの性能を低下させることが指摘され、それを解決する方法も提案されている（例えば、特許文献1参照。）。

【0005】

また、レプリケーションは、二相コミットにかかる時間が実用レベルを超えて遅いために開発された技術である。主にマスタテーブルを、新規トランザクションデータを入力するサーバに複製し、読み取り専用のテーブルとして扱う。従来のネットワーク環境では、特にISDNやフレームリレー方式で実装されるWAN上では伝送速度が高くないため、オリジナルテーブル上のデータの更新の都度リアルタイムに複製を更新することが現実的でない。このため、キャッシュのあるサーバからの更新情報を定期的に見に行き更新することが行なわれるので、オリジナルテーブルと複製の内容が同期するまで数分を要していた。このため使用範囲が限定されていた。

【0006】

一方、主メモリとして通常使用されるRAMは、通電が止まると記憶内容を失うが、比較的高速なデータの入出力が可能である。このため、プログラムのロードのための領域や一時記憶領域として使用されている。従来の商品化されたデータベース管理システムにおいては、その開発初期段階において、RAMは高価なもので、また、不揮発メモリは低速でより高価であったため、データを保管する記憶媒体としては、専ら停電しても記憶内容が失われない磁気ディスク装置が使われてきた。これは、その後継システムにも影響し、未だにデータベースの記憶装置として磁気ディスクを用いた装置が使用されている。

【0007】

従来のデータベースのバックアップは、バックアップ媒体として低速な記憶装置を想定しており、データベースを停止しないでバックアップを取るとバックアップ中に行なわれた更新とその直前までバックアップされた内容との間で整合性が取れなくなるため、データベースを停止させ、その瞬間のスナップショットを

バックアップ媒体に書き込む方法が用いられていた。

【0008】

さらに、SETI@home に代表される従来のグリッドコンピューティングは、参加者のネットワークに負担を掛けないプロセス分担型のみしか存在していない（例えば、非特許文献1 参照。）。これは、xDSL や FTTN、CATV 等のブロードバンドインターネットが普及する前のインターネットの環境においては、ISDN を使用した最大 128 Kbps の非常時接続型の環境下で、世界中に散在する多数のパーソナルコンピュータをインターネットで接続するためであった。プロセス分担型のグリッドコンピューティングにおいては、参加者はセンターからプロシージャとデータを受け取り、自分のコンピュータで自分の仕事を処理しながら、受け取ったジョブをバックグラウンドで計算処理し、結果をセンターに返すことが行なわれる。頻繁に新しいジョブが入り込み、その結果を返すというような参加者のネットワーク環境に負担となるような処理ではなく、一回のデータやプロシージャ等のネットワークからの入力、何時間単位の計算処理、結果のネットワークへの出力という、参加者のネットワーク環境に負担とならない処理のみが分担されるようになっている。

【0009】

【特許文献1】

特開 2001-306380 号公報（第2-3 頁）

【非特許文献1】

「今こそ知りたいグリッドコンピューティングのすべて」，日経 IT プロフェッショナル，株式会社日経 BP，2002 年 8 月号，p. 40-42

【発明が解決しようとする課題】

しかしながら、二相コミットやレプリケーションにおいては、あるコンピュータを分散データベースシステムに組み込むには、複雑な設定の手続きが必要である。そのため、データを分散することが容易に行なうことができないという課題がある。

【0010】

また、近年、例えば企業では社内 LAN が構築され、社員の机上では高性能な

パーソナルコンピュータが置かれ、多数の高性能なパーソナルコンピュータが社内LANに接続されている。しかし、これらのパーソナルコンピュータでは、ワードプロセッサやスプレッドシート処理プログラム、プレゼンテーション作成ツールなどが昼間だけ稼動している状態であり、CPU能力やメモリ容量、ディスク容量には余剰があり、有効に利用されていないという課題がある。

【0011】

また、このことは、企業内に限られることはなく、例えば、インターネットの常時接続の設備が提供された集合住宅においては、その各戸に設置されたパーソナルコンピュータのCPU能力やメモリ容量、ディスク容量は有効に利用されていない。

【0012】

また、データが分散された場合、データベースを停止させることが非常に困難となる。このため、従来のデータベースのバックアップの方法を使用することができないという課題もある。

【0013】

そこで、本発明は、データの分散が容易になり、ネットワークに接続されたコンピュータのCPU能力やメモリ容量、ディスク容量を有効に利用することができる分散型データベースシステムを提供することを目的とする。

【0014】

【課題を解決するための手段】

上記の課題を解決するために、本発明においては、データベースを管理するデータベース管理装置と、データベース管理装置のデータベースを管理するためのトポロジ管理サーバ装置と、トポロジ管理サーバ装置が管理する管理ドメイン内に配置された複数の計算機と、を備える分散型データベースシステムを提供する。この分散型データベースシステムにおいては、トポロジ管理サーバ装置は、データベース管理装置により管理されるデータベースオブジェクトを識別するためのデータベースオブジェクト識別子と、そのデータベースオブジェクトが管理されるデータベース管理装置を識別するためのデータベース管理装置識別子と、を対応付けた情報を含むトポロジ情報を格納する。また、トポロジ管理サーバは、

データベースオブジェクトをキャッシュするために計算機から発せられるキャッシュ命令を受信し、キャッシュ命令に含まれるデータベースオブジェクト識別子に基づいてデータベース管理装置識別子を取得し、そのデータベース管理装置識別子で識別されるデータベース管理装置に対して、キャッシュ要求を転送する。

【0 0 1 5】

これにより、計算機にデータベースオブジェクトをキャッシュさせることが容易に行なえ、課題が解決される。

【0 0 1 6】

また、計算機は、キャッシュしたデータベースオブジェクトを参照しながら計算を行なうためのプログラムを受信するようにしてもよい。

【0 0 1 7】

これにより、データを計算機に分散させて、計算を行なうことが可能となり、課題が解決される。

【0 0 1 8】

なお、計算機は、自身の計算負荷を検出し、その計算負荷に応じてキャッシュ要求を送信するようにしてもよい。

【0 0 1 9】

これにより、例えば、計算負荷が低い場合にデータベースオブジェクトのキャッシュが行なえ、課題が解決される。

【0 0 2 0】

また、キャッシュ要求には、キャッシュできるデータベースオブジェクトの量又は／及びデータ種別に関する情報が含まれていてもよい。

【0 0 2 1】

これにより、計算機のCPU能力やメモリ容量、ディスク容量に応じたキャッシュが行なえ、課題が解決される。

【0 0 2 2】

また、キャッシュ要求には、キャッシュが行なえる時間帯に関する情報が含まれていてもよい。

【0 0 2 3】

これにより、例えば、夜間にキャッシュが行なわれるようにでき、計算機の計算能力の余剰を効率良く利用でき、課題が解決される。

【0024】

また、キャッシュ要求には、計算機の計算能力に関する情報が含まれていてもよい。

【0025】

これにより、計算能力に応じたデータの分散が可能となり、課題が解決される。

【0026】

また、キャッシュ要求には、キャッシュに対する対価に関する情報が含まれていてもよい。

【0027】

これにより、例えば、計算機の所有者に本発明の分散型データベースシステムに参加させる動機付けを行なうことができる。

【0028】

また、トポロジ情報は、データベースオブジェクト識別子にデータベースオブジェクトがキャッシュされた計算機の識別子を対応付け、トポロジ管理サーバ装置は、データベースオブジェクトが計算機にキャッシュされることを検出すると、トポロジ情報を更新するようにしてもよい。

【0029】

これにより、データベースオブジェクトをキャッシュする計算機の追加が容易に行なえる。

【0030】

また、トポロジ管理サーバ装置は、データベースオブジェクトに対するロックの情報を保持するようにしてもよい。

【0031】

これにより、データが分散されても、矛盾を生じさせることなくデータの更新が行なえるようになる。

【0032】

また、トポロジ管理サーバ装置は、他のトポロジ管理サーバ装置とトポロジ情報を交換するようになっていてもよい。

【0033】

これにより、広範囲でのデータの分散が可能となる。

【0034】

また、データベース管理装置がデータベースオブジェクトの変更を行なうと、そのデータベースオブジェクトをキャッシュしている計算機に向けて変更を送信するようにしたり、あるいは、計算機がキャッシュしているデータベースオブジェクトの変更を行なうと、データベース管理装置に向けて変更を送信したりするようにしてもよい。

【0035】

これにより、データの変更を実現することができる。特に、計算機でデータベースオブジェクトを参照しながら計算が行なわれている場合には、データの変更に応じて計算を行なうことができる。

【0036】

また、データベース管理装置での変更操作をジャーナルとして送信し、それをジャーナル管理サーバ装置で受信し、ジャーナル管理サーバ装置で、ジャーナルをリプレイするようにしてもよい。

【0037】

これによりデータベースを停止させることなく、バックアップを行なうことができ、課題が解決される。

【0038】

【発明の実施の形態】

以下、本発明の実施の形態について、図を用いて説明する。なお、本発明は、これら実施の形態に何ら限定されるものではなく、その要旨を逸脱しない範囲において、種々なる態様で実施し得る。

【0039】

(発明の概念)

図1は、本発明の概念を示す。分散型データベースシステム101、113が

本発明に係る分散型データベースシステムである。例えば、分散型データベースシステム 101 は、データベース管理装置 102 と、トポロジ管理サーバ装置 103 と、複数の計算機 104、105、…、106 と、を備え、これらがネットワークシステムなどを構成するためのルータ 107 によって通信可能になっている。

【0040】

計算機 104、105、…、106 より、データベース管理装置 102 で管理されているデータベースオブジェクトをキャッシュするための要求であるキャッシュ要求が、トポロジ管理サーバ装置 103 へ送信される。トポロジ管理サーバ装置 103 はキャッシュ要求をデータベース管理装置 102 へ転送し、それに応じて、データベース管理装置 102 はデータベースオブジェクトを、キャッシュ要求を送信した計算機に向けて送信し、計算機がデータベースオブジェクトをキャッシュできるようになっている。

【0041】

また、図 1 に例示されるように、分散型データベースシステムが、複数存在してもよい。この場合、複数の分散型データベースシステムは、通信網 114 を介して接続されている。このような場合、分散型データベースシステム 101 のトポロジ管理サーバ装置 103 と、分散型データベースシステム 113 のトポロジ管理サーバ装置 109 と、が互いに通信をして、自分の属する分散型データベース管理システムのデータベース管理装置が格納するデータベースオブジェクトに関する情報を交換する。例えば、トポロジ管理サーバ装置 103 は、データベース管理装置 102 が格納するデータベースオブジェクトに関する情報を、トポロジ管理サーバ装置 109 へ送信する。

【0042】

これにより、例えば、分散型データベースシステム 113 の計算機 110 がトポロジ管理サーバ装置 109 へ、データベース管理装置 102 が管理するデータベースオブジェクトのキャッシュ要求を、送信することにより、トポロジ管理サーバ装置 109 は、要求されたデータベースオブジェクトが分散型データベースシステム 101 のデータベース管理装置 102 に存在することを検出し、キャッ

シユ要求をトポロジ管理サーバ装置 103 へ転送する。

【0043】

なお、トポロジ管理サーバ装置に対して、そのトポロジ管理サーバ装置へキャッシュ要求を送信する計算機が属する分散型データベースシステムを、「管理ドメイン」、あるいは、「トポロジドメイン」という場合がある。

【0044】

図2は、計算機104からキャッシュ要求が送信され、それがトポロジ管理サーバ装置に転送されて、データベース管理装置からデータベースオブジェクトの転送がされる過程を示すシーケンス図を例示する。

【0045】

ステップS201において、計算機104がキャッシュ要求をトポロジ管理サーバ装置103へ送信する。もし、何らかの理由により、データベースオブジェクトのキャッシュが許可できない場合には、キャッシュできない旨の返答が行なわれる(ステップS202)。

【0046】

もしキャッシュが可能であり、そのキャッシュの要求が他の管理ドメインのデータベース管理装置に格納されている場合には、キャッシュ要求をルータ107へ送信し(ステップS203)、ルータ、トポロジ管理サーバ装置109へ転送される(ステップS204、S205)。もし、何らかの理由により、データベースオブジェクトのキャッシュが許可できない場合には、キャッシュできない旨の返答が行なわれ、計算機104へ送信される(ステップS206、S207、S208)。

【0047】

もし、キャッシュが可能であれば、データベース管理装置108へキャッシュ要求が転送され(ステップS209)、データベース管理装置108よりキャッシュが送信され(ステップS210)、計算機104へ転送される(ステップS211、S212、S213)。

【0048】

また、キャッシュ要求が、同じ管理ドメインのデータベース管理装置102に

格納されたデータベースオブジェクトに対するものであれば、トポロジ管理サーバ装置 103 は、データベース管理装置 102 へキャッシュ要求を転送する（ステップ S 214）。データベース管理装置 102 は、キャッシュを送信し（ステップ S 215）、トポロジ管理サーバ装置 103 は、キャッシュを計算機 104 へ転送する（ステップ S 216）。

【0049】

もし、計算機 104 がデータベースオブジェクトのキャッシュを解放する場合には、トポロジ管理サーバ装置 103 へキャッシュ解放要求を送信する（ステップ S 217）。もし、キャッシュ解放要求が同じ管理ドメインのデータベース管理装置に格納されているデータベースオブジェクトに関するものであれば、トポロジ管理サーバ装置 103 は、キャッシュ解放完了通知を返す（ステップ S 216）。

【0050】

もし、キャッシュ解放要求が異なる管理ドメインのデータベース管理装置に格納されているデータベースオブジェクトに関するものであれば、トポロジ管理サーバ装置 103 は、異なる管理ドメインのトポロジ管理サーバ装置 108 へキャッシュ解放要求を送信し（ステップ S 219、S 220、S 221）、異なる管理ドメインのトポロジ管理サーバ装置 108 は、キャッシュ解放完了要求を返信する（ステップ S 222、S 223、S 224）。

【0051】

なお、ステップ S 215、S 210 において、キャッシュの送信は、データベース管理装置から同じ管理ドメインのトポロジ管理サーバ装置へ行なわれているが、キャッシュ要求を送信した計算機へ直接、キャッシュの送信が行なわれるようになっていてもよい。

【0052】

また、トポロジ管理サーバ装置は、データベースオブジェクトに対するロック操作を管理してもよい。

【0053】

図 3 は、計算機 104 が、キャッシュしているデータベースオブジェクトにロ

ックを行なう際の過程に関するシーケンス図を例示する。

【0054】

ステップS301において、計算機104がトポロジ管理サーバ装置103にSLOCK要求を送信したとする。もし、ロックの要求が同じ管理ドメインのデータベース管理装置が格納しているデータベースオブジェクトに対するものであり、ロックができない場合には、SLOCKができない旨の返答がトポロジ管理サーバ装置103より行なわれる（ステップS302）。また、ロックの要求が同じ管理ドメインのデータベース管理装置が格納しているデータベースオブジェクトに対するものであり、ロックが可能であれば、ロックができた旨の返答がされる（ステップS303）。

【0055】

もし、ロックの要求が異なる管理ドメインのデータベース管理装置が格納しているデータベースオブジェクトに対するものであれば、トポロジ管理サーバ装置103は、異なる管理ドメインのトポロジ管理サーバ装置109へ向けてロック要求を送信し（ステップS304、S305、S306）、もしロックができなければその旨の返答がされる（ステップS307、S308、S309、S310）。また、もし、ロックが可能であれば、ロックができた旨の返答がされる（ステップS311、S312、S313、S314）。

【0056】

このようにトポロジ管理サーバ装置が、管理ドメインのデータベース管理装置に格納されたデータベースオブジェクトのキャッシュやロックなどを管理することにより、新たな計算機を管理ドメインに追加したり、他の管理ドメインのデータベース管理装置が格納するデータベースオブジェクトのキャッシュを行なったりすることが可能となる。また、データの不整合が起きないように、データベースオブジェクトに対してロックを行なうことも可能となる。

【0057】

（実施形態1）

図4は、本発明の実施形態1における分散型データベースシステムの機能ブロック図を例示する。本実施形態における分散型データベースシステム400は、

データベース管理装置 402 と、トポロジ管理サーバ装置 401 と、複数の計算機 403、404、…、405 と、を備える。

【0058】

(実施形態 1：分散型データベースシステムの構成)

「データベース管理装置」402 は、ネットワーク上に配置されたデータベースを管理する。なお、ネットワーク上に配置されたデータベースには、データベース管理装置 402 の中に格納されているデータベースも含まれる。

【0059】

「トポロジ管理サーバ装置」401 は、データベース管理装置 402 のデータを分散するための装置であり、以下に説明する部を有する。

【0060】

「計算機」403、404、…、405 は、トポロジ管理サーバ装置 401 が前記ネットワーク上で管理する管理ドメイン内に配置されており、以下に説明する部を有する。なお、「前記ネットワーク」とは、データベース管理装置 402 が管理するデータベースが配置されたネットワークを意味する。

【0061】

(実施形態 1：トポロジ管理サーバ装置)

図 5 は、トポロジ管理サーバ装置 401 の機能ブロック図を例示する。トポロジ管理サーバ装置 401 は、トポロジ情報格納部 501 と、キャッシュ要求受信部 502 と、データベース管理装置識別子取得部 503 と、キャッシュ要求転送部 504 と、を有する。

【0062】

(実施形態 1：トポロジ管理サーバ装置：トポロジ情報格納部)

「トポロジ情報格納部」501 は、トポロジ情報を格納する。「トポロジ情報」とは、データベースオブジェクト識別子と、データベース管理装置識別子と、を対応付けた情報を含む情報である。「データベースオブジェクト識別子」とは、データベース管理装置 402 により管理されるデータベースオブジェクトを識別するための情報である。「データベースオブジェクト」には、例えば次のものがある；(1) データベースそのもの、(2) データベースを構成する個々のテ

ーブル、(3) テーブルの列に付けられた索引、(4) テーブルを構成する個々の行、(5) 行を構成する個々の列。したがって、データベースオブジェクト識別子は、データベース識別子、テーブル識別子、索引識別子、行識別子、列識別子などである。「データベース管理装置識別子」とは、データベースオブジェクトが管理されるデータベース管理装置を識別するための情報である。例えば、データベース管理装置が、名前で識別される場合には、その名前がデータベース管理装置識別子となり、また、例えば、IPアドレスで識別される場合は、そのIPアドレスがデータベース管理装置識別子となる。

【0063】

トポロジ情報は、データベースオブジェクト識別子と、データベース管理装置識別子と、を対応付けた情報を含む。そこで、トポロジ情報格納部501は、例えば、データベースオブジェクト識別子とデータベース管理装置識別子とからなる列を持つ表によりトポロジ情報を格納するようにしてもよい。データベースオブジェクト識別子から、それに対応付けられたデータベース管理装置識別子が得られるように、データベースオブジェクト識別子をキーとし、データベース管理装置識別子をバリューとする索引を用いるようにしてもよい。

【0064】

(実施形態1：トポロジ管理サーバ装置：キャッシュ要求受信部)

「キャッシュ要求受信部」502は、キャッシュ要求を受信する。「キャッシュ要求」とは、データベースオブジェクト識別子で識別されるデータベースオブジェクトをキャッシュするために、計算機の少なくとも一以上から発せられるデータベースオブジェクト識別子を含む情報である。

【0065】

図6は、キャッシュ要求を模式的に例示する。キャッシュ要求506が、データベースオブジェクト識別子601を含んでいる状態が示されている。

【0066】

(実施形態1：トポロジ管理サーバ装置：データベース管理装置識別子取得部)

「データベース管理装置識別子取得部」503は、キャッシュ要求受信部50

2で受信されたキャッシュ要求に含まれるデータベースオブジェクト識別子に基づいて、トポロジ情報格納部501から、対応するデータベース管理装置識別子を取得する。例えば、データベースオブジェクト識別子をキーとし、データベース管理装置識別子をバリューとする索引がある場合には、その索引を用いて、データベース管理装置識別子を取得する。

【0067】

(実施形態1：トポロジ管理サーバ装置：キャッシュ要求転送部)

「キャッシュ要求転送部」504は、データベース管理装置識別子取得部503で取得されたデータベース管理装置識別子で識別されるデータベース管理装置に対して、前記キャッシュ要求を転送する。「前記キャッシュ要求」とは、キャッシュ要求受信部で受信されたキャッシュ要求である。

【0068】

(実施形態1：トポロジ管理サーバ装置の処理)

図8は、トポロジ管理サーバ装置401の処理の流れを説明するフローチャートを例示する。ステップS801において、キャッシュ要求受信部502にて、キャッシュ要求を受信する。ステップS802において、キャッシュ要求に含まれるデータベースオブジェクト識別子を選別し、データベース管理装置識別子取得部503にて、データベース管理装置識別子を取得する。ステップS804において、データベース管理装置識別子で識別されるデータベース管理装置へ向けて、キャッシュ要求転送部504にて、キャッシュ要求を転送する。

【0069】

(実施形態1：トポロジ管理サーバ装置が計算機とデータベース管理装置との間の通信を中継しても良い)

なお、トポロジ管理サーバ装置は、ネットワークの通信におけるルータあるいはファイアウォールのように計算機とデータベース管理装置との間の通信を中継するものであってもよい。

【0070】

図9は、トポロジ管理サーバ装置401が、計算機403、404、…、405とデータベース管理装置402との間の通信を中継する場合の分散型データベ

ースシステム 900 の機能ブロック図を例示する。図 9 に例示される構成が採られる場合には、計算機から送信されたキャッシュ要求に対してデータベース管理装置 402 から送信されるデータベースオブジェクトは、トポロジ管理サーバ装置 401 により中継されて計算機へ送信されることになる。このため、トポロジ管理サーバ装置 401 は、どの計算機がどのデータベースオブジェクトに対するキャッシュ要求を送信したかを管理する必要がある。

【0071】

図 10 は、計算機がどのデータベースオブジェクトに対するキャッシュ要求を送信したかを管理するために用いられる表を例示している。「キャッシュ要求送信元」の列には、キャッシュ要求を送信した計算機を識別する識別子が格納される。図 10 では、IP アドレスが識別子として用いられている。「データベースオブジェクト識別子」の列には、計算機が送信したキャッシュ要求に含まれるデータベースオブジェクト識別子が格納される。トポロジ管理サーバ装置 401 が、データベース管理装置 402 より送信されたデータベースオブジェクトを受信すると、表 1001 を参照し、データベースオブジェクト識別子からキャッシュ要求送信元の計算機の識別子を得て、その識別子により識別される計算機へデータベースオブジェクトを送信する。

【0072】

(実施形態 1：計算機)

図 11 は、計算機の機能ブロック図を例示する。計算機 403 は、キャッシュ要求送信部 1101 と、データベースオブジェクト受信部 1102 と、データベースオブジェクトキャッシュ部 1103 と、を有している。

【0073】

(実施形態 1：計算機：キャッシュ要求受信部)

「キャッシュ要求受信部」1101 は、キャッシュ要求を送信する。キャッシュ要求を送信するタイミングとしては、例えば、計算機 403 の操作者が指令した時を挙げることができる。この操作者の指令の中に、キャッシュしようとするデータベースオブジェクトを識別するデータベースオブジェクト識別子が含まれていてもよい。あるいは、計算機 403 に対してキャッシュすべきデータベー

スオブジェクトを識別するデータベースオブジェクト識別子が前もって指定されていてもよい。

【0074】

(実施形態1：計算機：データベースオブジェクト受信部)

「データベースオブジェクト受信部」1102は、キャッシュ要求送信部1101でのキャッシュ要求の送信に応じて返信されるデータベースオブジェクトを受信する。

【0075】

(実施形態1：計算機：データベースオブジェクトキャッシュ部)

「データベースオブジェクトキャッシュ部」1103は、データベースオブジェクト受信部1102で受信されたデータベースオブジェクトをキャッシュする。このキャッシュは、計算機403の有する記憶装置によって行なわれる。例えば、ランダムアクセスメモリなどのメモリにより行なわれたり、あるいは、磁気ディスクなどの二次記憶により行なわれたりする。

【0076】

(実施形態1：計算機の処理)

図12は、計算機403の処理の流れを説明するフローチャートを例示する。ステップS1201において、キャッシュ要求送信部1101にて、キャッシュ要求を送信する。ステップS1102において、データベースオブジェクトの受信が可能となるまで待つ。ステップS1203において、データベースオブジェクト受信部1102にて、データベースオブジェクトを受信する。ステップS1204において、データベースオブジェクトキャッシュ部にて、データベースオブジェクトをキャッシュする。

【0077】

(実施形態1：データベース管理装置)

図13は、データベース管理装置402の機能ブロック図を例示する。データベース管理装置402は、キャッシュ要求受信部1301と、データベースオブジェクト複製送信部1302と、を有する。

【0078】

(実施形態1：データベース管理装置：キャッシュ要求受信部)

「キャッシュ要求受信部」1301は、トポロジ管理サーバ装置401から転送されるキャッシュ要求を受信する。

【0079】

(実施形態1：データベース管理装置：データベースオブジェクト複製送信部)

「データベースオブジェクト複製送信部」1302は、キャッシュ要求受信部で受信されたキャッシュ要求に応じてデータベースオブジェクトを複製して送信する。キャッシュ要求には、データベースオブジェクト識別子が含まれるので、そのデータベースオブジェクト識別子で識別されるデータベースオブジェクトを、データベース1303より取得し、複製して送信する。なお、図13では、データベース1303は、データベース管理装置402の内部に存在するが、データベース管理装置402の外部に存在していてもよい。例えば、データベース管理装置402以外のサーバによりデータベースが格納されていてもよい。

【0080】

(実施形態1：データベース管理装置の処理)

図14は、キャッシュ要求が受信されてからデータベースオブジェクトが送信されるまでのデータベース管理装置402の処理の流れを説明するフローチャートを例示する。ステップS1401において、キャッシュ要求受信部1301により、キャッシュ要求を受信する。ステップS1402において、キャッシュ要求に含まれるデータベースオブジェクト識別子を得る。ステップS1403において、データベースオブジェクトを得る。ステップS1404において、キャッシュ要求の送信元を得る。ステップS1405において、データベースオブジェクト複製送信部1302にて、キャッシュ要求の送信元へ、データベースオブジェクトの複製し送信する。

【0081】

なお、ステップS1404において、キャッシュ要求の送信元を得ることが行なわれるが、トポロジ管理サーバ装置401がキャッシュ要求の送信元として得られる場合と、キャッシュ要求を送信した計算機がキャッシュ要求の送信元とし

て得られる場合がある。また、キャッシュ要求の送信元を明示的に得ることを行なわなくても、キャッシュ要求の受信に伴い、自動的にキャッシュ要求の送信元が得られる場合もある。例えば、TCP/IPにおける通信においては、通信のためのコネクションを確立することにより、双方向の通信路が確立される場合がある。このような場合には、キャッシュ要求の受信により、データベースオブジェクトを送信するための通信路も確立されることになり、明示的なキャッシュ要求の送信元を得ることは不要となる。

【0082】

(実施形態1：計算機と、トポロジ管理サーバ装置と、データベース管理装置と、の間のデータの流れ)

図7は、計算機403よりキャッシュ要求が送信されとした場合の、計算機403と、トポロジ管理サーバ装置401と、データベース管理装置402と、の間の処理の流れを説明するシーケンス図を例示する。ステップS701において、計算機403よりキャッシュ要求が送信され、トポロジ管理サーバ装置401により受信される。ステップ702において、トポロジ管理サーバ装置401は、受信したキャッシュ要求からデータベースオブジェクト識別子を取得し、そのデータベースオブジェクト識別子からデータベース管理装置識別子取得部503によりデータベース管理装置識別子を取得し、そのデータベース管理装置識別子により識別されるデータベース管理装置へキャッシュ要求を転送する。データベース管理装置402が転送されたキャッシュ要求を受信すると、キャッシュ要求に含まれるデータベースオブジェクト識別子を取得し、そのデータベースオブジェクト識別子で識別されるデータベースオブジェクトを計算機403に向けて送信する(ステップS703)。なお、データベース管理装置402は、データベースオブジェクトを計算機403に直接送信するようにしてもよいし、あるいは、トポロジ管理サーバ装置401を経由して計算機403に送信するようにしてもよい。

【0083】

(データベース管理装置、トポロジ管理サーバ装置、計算機は、機械的要素を含まないものであってもよい)

なお、データベース管理装置、トポロジ管理サーバ装置、計算機は、コンピュータ装置によって実現される。この場合、データベース管理装置、トポロジ管理サーバ装置、計算機を実現するコンピュータ装置のいずれか一以上あるいは全ては、回転軸などの摺動機械要素を含む磁気ディスク装置を使用しないものであってもよい。このように機械要素が存在しないようにすることにより、コンピュータ装置の信頼性を上げることができ、システム全体の信頼性をも上げることができる。また、磁気ディスクを使用しないことにより、コンピュータ装置で動作するオペレーティングシステムにファイルシステムを持たせる必要がなくなり、リソースの最大有効利用が可能となる。また、コンピュータ装置には蓄電池を装備し停電が発生してもある程度の時間の電源の供給が可能な無停電装置を接続しておいてもよい。これにより、さらに信頼性を上げることができる。

【0084】

(実施形態1：主な効果)

本実施形態の分散型データベースシステムにより、計算機がキャッシュ要求を送信し返信されるデータベースオブジェクトをキャッシュする機能を持てば、分散型データベースシステムに参加することができ、データの分散が容易に行なえる。また、データベースオブジェクトの複製が計算機によりキャッシュされるので、データベース管理装置に障害が発生しても、計算機のキャッシュを用いて、復旧を行なうことができる。また、データベースオブジェクトを計算機にキャッシュさせることにより、データベースオブジェクトの多重化（物理的に別の場所に存在する複数のコンピュータ装置へのミラーリング）が可能となる。このような多重化により、データベース管理装置を点検や交換などにより停止せざるを得ないじょうきょうにおいて、システム全体のサービスを停止することなく継続することが可能となる。

【0085】

(実施形態2)

本発明の実施形態2においては、計算機にキャッシュされたデータベースオブジェクトを参照しながら計算が行なわれる分散型データベースシステムが提供される。すなわち、グリッドコンピューティングが実現できる。

【0086】

(実施形態2：構成)

図15は、実施形態2における分散型データベースシステムの計算機403の機能ブロック図を例示する。本実施形態の計算機は、実施形態1の計算機が更にプログラム受信部1501を有したものになっている。

【0087】

(実施形態2：計算機：プログラム受信部)

「プログラム受信部」1501は、データベースオブジェクトキャッシュ部1103にキャッシュされたデータベースオブジェクトを参照しながら計算を行なうためのプログラムを受信する。例えば、ftpなどを用いて、計算機が外部のサーバなどとの通信を開始してプログラムを受信してもよい。また、外部のサーバなどが、計算機403へ定期的あるいは不定期的にプログラムを送信し、それを受信するようにしてもよい。なお、外部のサーバなどは、分散型データベースシステムの内部にあってもよいし、外部にあってもよい。また、プログラムが光ディスクなどの媒体に記録され、その媒体を読み取ることにより、プログラムが受信されてもよい。

【0088】

プログラムは、データベースオブジェクトキャッシュ部1103にキャッシュされたデータベースオブジェクトを参照しながら計算を行なうものであれば、何であってよい。例えば、有限要素法の各要素がデータベースオブジェクトにより表現され、そのようなデータベースオブジェクトにより、有限要素法の計算を行なうものであってもよい。また、計算の結果をデータベースオブジェクトに書き込むことにより、他の計算機で動作するプログラムに、計算の結果を伝えるようになっていてもよい。このようなプログラムにより行なわれる計算の例としては、例えば、海洋での汚染物質の拡散の様子を計算するもの、気象予報を行なうもの、などを挙げることができる。

【0089】

(実施形態2：主な効果)

本実施形態の分散型データベースシステムにより、データベースオブジェクト

を参照しながら行なわれる計算を分散して行なうことができる。特に、データベースオブジェクトを更新しながら計算を行なうことにより、動的に変化する状況に応じた計算を行なうことができる。

【0090】

(実施形態3)

本発明の実施形態3においては、負荷に基づいてキャッシュ要求を送信する計算機を備える分散型データベースシステムが提供される。

【0091】

(実施形態3：構成)

図16は、本実施形態における分散型データベースシステムの計算機の機能ブロック図を例示する。本実施形態における計算機403は、実施形態1または2の計算機が、計算負荷検出部1601を更に有している構成になっている。

【0092】

(実施形態3：計算機：計算負荷検出部)

「計算負荷検出部」1601は、自身の計算負荷を検出する。ここに「自身」とは、計算機403を意味する。また、「計算負荷」とは、計算機403のCPU占有率、メモリ占有率、磁気ディスクのデータの占有率、通信の状況（例えば、破棄されたパケットの数、一定時間あたりに送信あるいは受信されたパケットの数）、などである。

【0093】

(実施形態3：計算機：キャッシュ要求送信部)

本実施形態において、キャッシュ要求送信部1101は、計算負荷検出部1601で検出された計算負荷に基づいてキャッシュ要求を送信する。例えば、メモリ占有率が所定の値以下であれば、キャッシュ要求送信部1101は、キャッシュ要求を送信する。あるいは、CPU占有率が所定の値以下であれば、キャッシュ要求を送信する。

【0094】

また、計算負荷検出部1601で検出された計算負荷に応じてキャッシュ要求の内容が変更されてもよい。例えば、メモリ占有率が低ければ、サイズの大きな

データベースオブジェクトを識別するデータベースオブジェクト識別子を含むキャッシュ要求が送信されるようになっていてもよい。

【0095】

(実施形態3：主な効果)

本実施形態によれば、計算負荷に基づいてキャッシュ要求が送信されるので、例えば、計算負荷の高い場合には、キャッシュ要求が送信されないで、計算機の計算負荷を異常に高めることを避けることができる。

【0096】

(実施形態4)

本実施形態においては、起動時にキャッシュ要求が送信される計算機を備える分散型データベース管理システムが提供される。

【0097】

(実施形態4：計算機：キャッシュ送信部)

本実施形態における計算機のキャッシュ要求送信部は、実施形態1または2の分散型データベースシステムが備える計算機が起動すると、キャッシュ要求を送信することを特徴としたものである。ここに、「計算機」とは、キャッシュ要求送信部を有する計算機であってもよいし、他の計算機であってもよい。例えば、計算機が起動し、ブートする過程において、キャッシュ要求送信部が活性化されて、キャッシュ要求が送信されるようになっていてもよい。また、他の計算機が起動すると、その情報が受信されてキャッシュ要求送信部が活性化されてキャッシュ要求が送信されるようになっていてもよい。

【0098】

(実施形態4：主な効果)

本実施形態により、計算機が起動するとキャッシュ要求が送信されるので、キャッシュ要求を送信する操作を別に行なう必要がない。これにより、多数の計算機がキャッシュ要求を送信できるようになり、データの分散を図ることができる。

【0099】

(実施形態5)

本実施形態においては、キャッシュ要求に計算機がキャッシュすることができるデータベースオブジェクトの量又は／及びデータ種別に関する情報が含まれる分散型データベースシステムが提供される。

【0100】

(実施形態5：キャッシュ要求)

本実施形態は、実施形態1ないし4のいずれかの分散型データベースシステムにおいて、キャッシュ要求には、データベースオブジェクトキャッシュ部がキャッシュすることができるデータベースオブジェクトの量又は／及びデータ種別に関する情報が含まれることを特徴とする。

【0101】

図17(a)は、キャッシュ要求に、データベースオブジェクトキャッシュ部がキャッシュすることができるデータベースオブジェクトの量に関する情報1701が含まれていることを模式的に示す図である。例えば、どの程度のバイト数までのデータベースオブジェクトをキャッシュすることができるかの情報がキャッシュ要求に含まれている。

【0102】

図17(b)は、キャッシュ要求に、データベースオブジェクトキャッシュ部がキャッシュすることができるデータ種別に関する情報1702が含まれていることを模式的に示す図である。データ種別としては、例えば、データベース、テーブル、索引、行、列などがある。

【0103】

本実施形態において、データベースオブジェクト識別子とデータベースオブジェクトの量とがキャッシュ要求に含まれるとき、そのキャッシュ要求は、例えば、データベースオブジェクト識別子がデータベースを識別するものである場合には、データベースがデータベースオブジェクトの量と等しいかあるいはそれより小さい場合にキャッシュすることを要求する条件付きのキャッシュ要求と解釈してもよい。また、データベースの中のテーブルのうち、データベースオブジェクトの量と等しいかあるいはそれより小さいものを要求するキャッシュ要求と解釈してもよい。データベースの中のテーブルのうち、データベースオブジェクトの

量と等しいかあるいはそれより小さいものの選択をデータベース管理装置に行なわせることにより、特定のテーブルだけがキャッシュされることを防止することができる。

【0104】

同様に、データベースオブジェクト識別子とデータ種別とがキャッシュ要求に含まれる場合には、そのキャッシュ要求は、例えば、データベースオブジェクト識別子がデータベースを識別し、データ種別がテーブルを指定するものであるときには、データベースオブジェクト識別子により識別されるデータベースに含まれるテーブルを要求するキャッシュ要求と解釈してもよい。

【0105】

(実施形態5：キャッシュ要求は、計算負荷に応じて送信されてもよい)

本実施形態において、キャッシュ要求は、例えば、計算負荷に応じて変更がされたり、あるいは、計算機の起動時に送信されたりするようになっていてもよい。例えば、メモリ占有率が低ければ、大きなサイズのデータベースオブジェクトをキャッシュするためのキャッシュ要求が送信されるようになっていてもよい。

【0106】

(実施形態5：計算により参照されるデータベースオブジェクトに基づいてキャッシュ要求が送信されてもよい)

また、データベースオブジェクトを参照して計算を行なうプログラムが計算機で動作される場合には、そのプログラムの参照するデータベースオブジェクトに応じてキャッシュ要求が送信されてもよい。例えば、同じテーブルの行が多数参照される場合には、それらの行を含むテーブルを識別するデータベースオブジェクト識別子を含むキャッシュ要求が送信されるようになっていてもよい。また、多数のテーブルの行が参照される計算が行なわれる場合には、それらのテーブルを含むデータベースを識別するデータベースオブジェクト識別子を含むキャッシュ要求が送信されるようになっていてもよい。

【0107】

(実施形態5：キャッシュ要求には、データベースオブジェクトの量に関する情報とデータ種別との両方が含まれていてもよい)

なお、キャッシュ要求には、データベースオブジェクトの量に関する情報とデータ種別との両方が同時に含まれていてもよい。これにより、特定の種別のデータベースオブジェクトのうち、特定のサイズまでのもののキャッシュを要求することができる。

【0108】

(実施形態5：主な効果)

本実施形態により、計算機がキャッシュすることになるデータベースオブジェクトが過大なものになることを防止することができる。また、計算により参照されるデータベースオブジェクトを効率良く参照するためのキャッシュを行なうことができる。また、計算機にキャッシュされるデータベースオブジェクトが偏らないようにすることができる。

【0109】

(実施形態6)

本実施形態においては、キャッシュ要求に、キャッシュが行なえる時間帯に関する情報が含まれる分散型データベースシステムが提供される。

【0110】

(実施形態6：キャッシュ要求)

本実施形態は、実施形態1ないし4のいずれかの分散型データベースシステムにおいて、キャッシュ要求には、データベースオブジェクトキャッシュ部がデータベースオブジェクトをキャッシュできる時間帯に関する情報が含まれる。

【0111】

図18は、キャッシュ要求にデータベースオブジェクトキャッシュ部がデータベースオブジェクトをキャッシュできる時間帯に関する情報1801が含まれていることを模式的に示す図である。

【0112】

時間帯に関する情報1801は、例えば、データベースオブジェクトキャッシュ部がデータベースオブジェクトをキャッシュできる時間帯がいつであることを示すものであってもよい。また、データベースオブジェクトキャッシュ部がデータベースオブジェクトをキャッシュできない時間帯がいつであることを示すものであ

ってもよい。

【0113】

このような時間帯に関する情報は、計算機の稼動予定に基づいて生成されたり、あるいは、過去の計算機の稼動状況に基づいて生成されたりしてもよい。例えば、昼間に計算機が稼動しない日が過去に多ければ、昼間はキャッシュすることができないという時間帯に関する情報が生成されるようになっていてもよい。

【0114】

また、過去の計算機の計算負荷から時間帯に関する情報が生成されてもよい。例えば、夕方は、CPU占有率が高かったり、通信パケット数が多い場合には、夕方にはキャッシュすることができないという時間帯に関する情報が生成されるようになっていてもよい。

【0115】

また、データベースオブジェクトを参照しながら計算を行なうプログラムがいつ動作するかに基づいて、キャッシュできる時間帯に関する情報が生成されるようになっていてもよい。

【0116】

(実施形態6：主な効果)

本実施形態により、データベースオブジェクトキャッシュ部がキャッシュできる時間帯に、データベース管理装置が計算機に対してデータベースオブジェクトを送信することができ、キャッシュを有効に行なうことができる。

【0117】

(実施形態7)

本実施形態においては、キャッシュ要求に、計算機の有する情報が含まれる分散型データベースシステムが提供される。

【0118】

(実施形態7：キャッシュ要求)

本実施形態は、実施形態1ないし4のいずれかの分散型データベースシステムにおいて、キャッシュ要求には、計算機の有する計算能力に関する情報が含まれることを特徴とする。

【0119】

図19は、キャッシュ要求に計算能力に関する情報1901が含まれていることを模式的に示す図である。

【0120】

「計算機の有する計算能力に関する情報」とは、例えば、計算機が備えるメモリ量、磁気ディスクの容量、CPUの型番や動作周波数など、計算の処理能力を示す情報である。

【0121】

(実施形態7：主な効果)

本実施形態によれば、計算能力に応じて、データベースオブジェクトのキャッシュを計算機にさせることができ、データの分散を有効に行なうことができる。

【0122】

(実施形態8)

本実施形態においては、キャッシュ要求にデータベースオブジェクトをキャッシュすることに対する対価に関する情報が含まれる分散型データベースシステムが提供される。

【0123】

(実施形態8：キャッシュ要求)

本実施形態は、実施形態1ないし4のいずれかの分散型データベースシステムにおいて、キャッシュ要求には、計算機がデータベースオブジェクトをキャッシュすることに対する対価に関する情報が含まれる。

【0124】

図20は、キャッシュ要求に、対価に関する情報2001が含まれていることを模式的に示す図である。

【0125】

「対価に関する情報」としては、例えば、一つのデータベースオブジェクトあたり、いくらの対価を計算機の管理者（所有者、占有者などを含む。）に支払うべきかを示す情報を挙げることができる。また、データベースオブジェクトのバイト数に応じていくらの対価を計算機の管理者に支払うべきかを示す情報を挙げ

ることができる。

【0126】

また、逆に、計算機の管理者がデータベース管理装置の管理者にいくらの対価を支払うべきかを示す情報であってもよい。この場合は、例えば、計算機で、データベースオブジェクトを参照して計算が行なわれる場合に、そのデータベースオブジェクトを参照することに対する対価となる。この場合、キャッシュ要求に含まれる対価に関する情報は、計算機の管理者が支払ってもよい対価の上限を示すと解釈されてもよい。

【0127】

(実施形態8：主な効果)

本実施形態によれば、例えば、データベース管理装置の管理者がデータベースオブジェクトを計算機にキャッシュさせてデータベースのバックアップを行なう場合において、最も費用が低くなるように計算機にキャッシュをさせることができる。また、データベースオブジェクトを参照して計算を行なう場合に、データベースオブジェクトの参照に伴って対価を支払う場合には、支払うべき対価が高くなることを防止することができる。

【0128】

(実施形態9)

本実施形態においては、データベースオブジェクトをキャッシュしている計算機を管理することができる分散型データベースシステムが提供される。

【0129】

(実施形態9：構成)

本実施形態の分散型データベースシステムは、実施形態1の分散型データベースシステムにおいて、トポロジ情報が、さらにデータベースオブジェクト識別子に、データベースオブジェクトがキャッシュされた計算機の識別子を対応付けたものである。

【0130】

また、本実施形態において、トポロジ管理サーバは、キャッシュ済情報受信部と、トポロジ情報キャッシュ更新部とを有する。

【0131】

(実施形態9：トポロジ情報)

図21は、本実施形態におけるトポロジ情報を模式的に例示する図である。トポロジ情報505に、データベースオブジェクト識別子2101にデータベース管理装置識別子2102とが対応付けられるとともに、計算機の識別子2103、2104が対応づけられている。この計算機の識別子2103、2104は、データベースオブジェクト識別子で識別されるデータベースオブジェクトをキャッシュしている計算機を識別する識別子である。その識別子としては、例えば、計算機のIPアドレスを用いることができる。

【0132】

図22は、図21に模式的に例示されたトポロジ情報を表現する表を例示する。表2201は、データベースオブジェクト識別子とデータベース管理装置識別子とを対応付ける表である。また、表2202は、データベースオブジェクト識別子と計算機の識別子とを対応付ける表である。このように二つの表を用いることにより、一つのデータベースオブジェクトが複数の計算機によりキャッシュされる場合には、表2202に複数のデータベースオブジェクト識別子と計算機の識別子とが格納され、表2201において、データベース管理装置識別子が冗長に格納されることを避けることができる。

【0133】

(実施形態9：トポロジ管理サーバ装置)

図24は、本実施形態におけるトポロジ管理サーバ装置の機能ブロック図を例示している。本実施形態におけるトポロジ管理サーバ装置は、実施形態1の分散型データベースシステムのトポロジ管理サーバ装置が、キャッシュ済情報受信部2401と、トポロジ情報キャッシュ更新部2402と、をさらに有している構成となっている。

【0134】

(実施形態9：トポロジ管理サーバ装置：キャッシュ済情報受信部)

「キャッシュ済情報受信部」2401は、キャッシュ済情報を受信する。「キャッシュ済情報」は、データベースオブジェクトが計算機にキャッシュされたこ

とを示す情報である。このキャッシュ済情報は、例えば、計算機にキャッシュ済情報送信部があり、計算機がデータベースオブジェクトをキャッシュすると、キャッシュ済情報送信部が、キャッシュ済情報を送信するようになっていてもよい。あるいは、データベース管理装置がキャッシュ済情報送信部を有しており、データベースオブジェクトを計算機に向けて送信することが終了した時点で、キャッシュ済情報を送信するようになっていてもよい。あるいは、データベースオブジェクトがトポロジ管理サーバ装置を経由して計算機に送信される場合には、トポロジ管理サーバ装置が計算機へデータベースオブジェクトを送信した後で、トポロジ管理サーバ装置内部でキャッシュ済情報が発生され、キャッシュ済情報受信部 2401 で受信されるようになっていてもよい。

【0135】

キャッシュ済情報には、データベースオブジェクトをキャッシュした計算機を識別するための識別子が少なくとも含まれる。また、更に、キャッシュしたデータベースオブジェクトのデータベースオブジェクト識別子が、キャッシュ済情報に、含まれていてもよい。

【0136】

(実施形態 9：トポロジ管理サーバ装置：トポロジ情報キャッシュ更新部)

「トポロジ情報キャッシュ更新部」2402 は、キャッシュ済情報受信部 2401 で受信されたキャッシュ済情報に基づいて、トポロジ情報格納部に格納されたトポロジ情報のキャッシュ済情報を最新の状態に更新する。

【0137】

例えば、キャッシュ済情報に、計算機の識別子とデータベースオブジェクト識別子とが含まれている場合には、データベースオブジェクト識別子と計算機の識別子とを取り出し、表 2202 に行を追加する。

【0138】

また、分散型データベースシステムが、図 9 に例示されるように、計算機とデータベース管理装置との通信がトポロジ管理サーバ装置を介して行なわれるように構成されている場合において、キャッシュ済情報に計算機の識別子しか含まれていないときには、例えば、図 10 の表 1001 を参照して、データベースオブ

ジェクト識別子を取り出し、表 2202 に行を追加するようにしてもよい。

【0139】

(実施形態 9：データの流れ)

図 25 は、計算機がキャッシュ済情報送信部を有している場合のデータの流れを例示するシーケンス図である。

【0140】

ステップ S2501 において、計算機 403 より、トポロジ管理サーバ装置 401 にキャッシュ要求が送信される。ステップ S2502 において、トポロジ管理サーバ装置 401 により、キャッシュ要求が、データベース管理装置 402 へ転送される。ステップ S2503 において、データベース管理装置 402 から計算機 403 へ向けてデータベースオブジェクトが送信される。ステップ S2504 において、計算機 403 がデータベースオブジェクトをキャッシュすると、キャッシュ済情報がトポロジ管理サーバ装置 401 へ送信される。

【0141】

(実施形態 9：主な効果)

本実施形態により、データベースオブジェクトをキャッシュしている計算機をトポロジ管理サーバ装置により管理することができる。これにより、例えば、データベース管理装置に障害が発生しても、データベースオブジェクトをキャッシュしている計算機を求めて、その計算機からキャッシュしているデータベースオブジェクトを取得して、障害からの復旧を行なうことができる。また、データベースオブジェクトのキャッシュに伴い、対価の支払が発生する場合に、その対価の支払についての情報をトポロジ管理サーバ装置で管理することができる。

【0142】

(実施形態 10)

本実施形態においては、データベースオブジェクトに関するロック情報を管理するトポロジ管理サーバ装置を備える分散型データベースシステムが提供される。

【0143】

(実施形態 10：構成)

本実施形態においては、実施形態 9 に記載の分散型データベースシステムにおいて、トポロジ情報が、さらにデータベースオブジェクト識別子に、ロック情報を対応付けることとする。また、トポロジ管理サーバ装置は、ロック操作情報受信部と、トポロジ情報ロック更新部と、をさらに有する。

【0144】

(実施形態 10：ロック情報)

「ロック情報」とは、データベースオブジェクトに操作されたロックに関する情報である。「ロック」とは、データベースに対する操作を一定の範囲で禁止することを意味する。例えば、データベースオブジェクトに対して読み取りの操作を行なう場合には、その読み取りの操作の間に、データベースオブジェクトの内容が変更されないように、共有ロック (S LOCK) をかけることを行なう。また、データベースオブジェクトの内容が変更される場合、その内容の変更が確定するまで (すなわち、コミットされるまで)、他の者によりデータベースオブジェクトの内容が読み取られたり、内容の変更が行なわれたりしないように排他ロック (X LOCK) をかけることを行なう。

【0145】

また、ロック情報には、以上説明したように、ロックの内容に限定されず、更に、ロック操作を行なった者が誰であるかを示す情報が含まれていてもよい。また、楽観的ロックを実現するために、ロックの操作が行なわれた時刻に関する情報も含まれていてもよい。

【0146】

また、ロック情報は、データベースオブジェクトに対してロックを行なったことを示す情報を含む場合に加えて、データベースオブジェクトに対して行なったロックを解除することを示す情報を含む場合があってもよい。

【0147】

本実施形態において、トポロジ情報にロック情報が含まれるために、トポロジ情報格納部は、トポロジ情報を、図 22 の表 2201、2202 に加えて、図 23 に例示される表 2301 を用いて管理するようにしてもよい。表 2301 は、データベースオブジェクト識別子とロック情報を対応付けている表である。

【0148】

(実施形態10：トポロジ管理サーバ装置)

図26は、本実施形態における分散型データベースシステムのトポロジ管理サーバ装置の機能ブロック図を例示している。トポロジ管理サーバ装置401は、実施形態9における分散型データベースシステムのトポロジ管理サーバ装置が、さらにロック操作情報受信部2601と、トポロジ情報ロック更新部2602と、を有している。

【0149】

(実施形態10：トポロジ管理サーバ装置：ロック操作情報受信部)

「ロック操作情報受信部」2601は、ロック情報を受信する。ロック情報は、例えば、データベース管理装置から送信されてもよい。また、計算機がキャッシュしているデータベースオブジェクトを参照したり変更したりする場合に、計算機がロック情報を送信してもよい。

【0150】

(実施形態10：トポロジ管理サーバ装置：トポロジ情報ロック更新部)

「トポロジ情報ロック更新部」2602は、ロック操作情報受信部2601で受信されたロック情報に基づいてトポロジ情報格納部501に格納されたトポロジ情報のロック情報を最新の状態に更新する。

【0151】

例えば、データベースオブジェクトに対してロックが行なわれる場合には、表2301に行を追加する。なお、この際、既にデータベースオブジェクトに対してロックが行なわれているかどうかをチェックし、もしロックが既に行なわれている場合に、新たに行なわれるロックと矛盾しないかどうかをチェックしてもよい。例えば、あるデータベースオブジェクトに対して、XLOCKが行なわれているときに、そのXLOCKを行なった者以外の者がそのデータベースオブジェクトにSLOCKを行なうと、ロックが矛盾するので、この場合は、トポロジ情報のロック情報を更新せず、エラーを返す。

【0152】

また、データベースオブジェクトのロックを解除する場合には、解除されるロ

ックに対応する表 2301 の対応する行を削除する。

【0153】

(実施形態 10：主な効果)

本実施形態によれば、ロックがトポロジ管理サーバ装置により管理されるので、データベースオブジェクトの更新を矛盾無く行なうことが可能となる。

【0154】

(実施形態 11)

本発明の実施形態 11 においては、複数のトポロジ管理サーバ装置がトポロジ情報を交換する分散型データベースシステムが提供される。

【0155】

図 27 は、本実施形態の概要を例示する。本実施形態の分散型データベースシステムは、実施形態 9 または 10 の分散型データベースシステムにおいて、トポロジ管理サーバ装置が複数ある。その複数のトポロジ管理サーバ装置は互いに通信が可能であって、トポロジ情報を交換しあう。このようなトポロジ情報の交換は、隣接するルータによる経路情報の交換に対応するものである。

【0156】

(実施形態 11：構成)

図 28 は、本実施形態における分散型データベースシステムのトポロジ管理サーバ装置の機能ブロック図を例示する。本実施形態のトポロジ管理サーバ装置は、実施形態 9 または 10 のトポロジ管理サーバ装置がトポロジ情報交換部 2901 を有するものになっている。

【0157】

(実施形態 11：トポロジ管理サーバ装置：トポロジ情報交換部)

「トポロジ情報交換部」2801 は、他トポロジ管理サーバ装置とトポロジ情報を交換する。「他トポロジ管理サーバ装置」とは、ネットワークにより通信が可能な他の管理ドメインを管理する他のトポロジ管理サーバ装置を意味する。

【0158】

トポロジ情報交換部 2801 は、定期的、あるいは、不定期的にトポロジ情報格納部 501 に格納されているトポロジ情報を送信する。また、他トポロジ管理

サーバ装置から、定期的、あるいは、不定期的に送信されるトポロジ情報を受信する。また、自身の存在を他トポロジ管理サーバ装置に知らせるために、トポロジ情報交換部は、自身のIPアドレスなどの識別子を含む情報をブロードキャストするようにしてもよい。ブロードキャストする情報には、例えば、トポロジ情報を送信する時間間隔や、格納しているトポロジ情報の量、などが含まれていてもよい。

【0159】

トポロジ情報交換部2801が他トポロジ管理サーバ装置よりトポロジ情報を受信した場合には、そのトポロジ情報をトポロジ情報格納部501に格納する。その際、トポロジ情報に、どの管理ドメインのトポロジ情報であるかを付加して格納してもよい。

【0160】

(実施形態11：主な効果)

本実施形態により、管理ドメインが複数ある場合に、計算機が属する管理ドメイン以外の管理ドメインのデータベース管理装置により管理されるデータベースオブジェクトをキャッシュすることが可能となる。これにより、広範囲のデータの分散が可能となる。

【0161】

(実施形態12、13)

本発明の実施形態12と13は、実施形態2における分散型データベースシステムにおいて、データベースオブジェクトの更新の機能を明確にしたものに関する。実施形態12は、データベース管理装置が、自身の管理するデータベースオブジェクトの更新を開始できるようにする形態であり、実施形態13は、計算機がキャッシュしているデータベースオブジェクトの更新を開始することができる形態である。実施形態13においては、計算機がキャッシュしているデータベースオブジェクトの更新は、例えば、プログラム受信部で受信されたプログラムの実行に伴って開始される。これにより、プログラムがデータベースオブジェクトの更新を行なうことができるようにしている。

【0162】

(実施形態 12)

図 29 は、実施形態 12 におけるデータの流れを説明するシーケンス図を例示する。データベース管理装置 402 でデータベースオブジェクトに対して更新操作が行なわれると、ステップ S2901 において、計算機によりキャッシュされているデータベースオブジェクトを更新操作の命令である更新操作命令が送信され、トポロジ管理サーバ装置 401 で受信される。ステップ S2902 において、トポロジ管理サーバ装置 401 は計算機に向けて、更新操作命令を計算機に向けて送信する。なお、トポロジ管理サーバ装置が格納するトポロジ情報がデータベースオブジェクト識別子にデータベースオブジェクトがキャッシュされた計算機の識別子に対応付けている場合には、更新されたデータベースオブジェクトをキャッシュしている計算機の識別子を求め、識別子が求められた計算機に対して更新操作命令が送信されるようになっていてもよい。

【0163】

なお、「更新操作」には、データベースオブジェクトの変更、削除が含まれる。また、データベースオブジェクトがデータベースであれば、データベースへのテーブルや索引の追加も含まれる。また、データベースオブジェクトがテーブルであれば、行の追加、列の追加を含み、また、列の名前の変更なども含まれる。

【0164】

(実施形態 12：データベース管理装置)

図 30 は、データベース管理装置の機能ブロック図を例示する。本実施形態のデータベース管理装置は、実施形態 2、実施形態 9 または実施形態 11 のデータベース管理装置が、更新操作命令送信部 3001 を有している。

【0165】

(実施形態 12：データベース管理装置：更新操作命令送信部)

「更新操作命令送信部」3001 は、データベースオブジェクトをキャッシュしている計算機であるクライアント装置に向けて、キャッシュされているデータベースオブジェクトの更新操作の命令である更新操作命令を送信する。例えば、全ての計算機に対してブロードキャストにより送信してもよい。もし、データベース管理装置が、データベースオブジェクトをキャッシュしている計算機を管理

していれば、更新操作命令送信部から直接計算機へ更新操作命令が送信されてよい。

【0166】

(実施形態12：実施形態9に基づく場合)

あるいは、本実施形態の分散型データベースシステムが実施形態9に基づく場合には、トポロジ管理サーバ装置が管理するトポロジ情報にどの計算機がどのデータベースオブジェクトをキャッシュしているかを保持している。そこで、更新操作命令送信部3001は、更新操作命令をトポロジ管理サーバ装置へ送信してもよい。更新操作命令を受信したトポロジ管理サーバ装置は、トポロジ情報に基づいて、更新されたデータベースオブジェクトをキャッシュしている計算機を求め、その計算機へ更新操作命令を送信する。

【0167】

(実施形態12：実施形態11に基づく場合)

また、本実施形態の分散型データベースシステムが実施形態11に基づく場合には、複数のトポロジ管理サーバ装置がトポロジ情報を交換している。したがって、他のドメインの計算機がデータベースオブジェクトをキャッシュしている場合であっても、他のドメインの計算機へ向けて更新操作命令を送信することができる。

【0168】

(実施形態12：計算機)

図31は、計算機の機能ブロック図を例示する。本実施形態の計算機は、実施形態2、実施形態9または実施形態11の計算機が、更新操作命令受信部3101と、データベースオブジェクト更新操作部3102を有している。

【0169】

(実施形態12：更新操作命令受信部)

「更新操作命令受信部」3101は、更新操作命令を受信する。

【0170】

(実施形態12：データベースオブジェクト更新操作部)

「データベースオブジェクト更新操作部」3102は、更新操作命令受信部3

101で受信された更新操作命令に基づいて、データベースオブジェクトキャッシュ部1103にキャッシュされたデータベースオブジェクトを更新する。

【0171】

例えば、更新操作命令が、データベースオブジェクトの削除を命令するものであれば、キャッシュしているデータベースオブジェクトの削除を行なう。

【0172】

また、更新操作命令に基づいて、データベースオブジェクトを更新する際には、更新操作命令の対象となるデータベースオブジェクトがデータベースオブジェクトキャッシュ部1103にキャッシュされているかどうかを調べ、キャッシュされている場合に限り、更新操作を行なうようにしてもよい。

【0173】

(実施形態12：主な効果)

本実施形態により、データベース管理装置でのデータベースオブジェクトに対する更新操作が、計算機にキャッシュされているデータベースオブジェクトに対して反映される。これにより、計算機で動作するプログラムは、最新の状態のデータベースオブジェクトを参照して計算を行なうことができる。

【0174】

また、本実施形態が実施形態11に基づく場合には、他ドメインの計算機に向けて更新操作命令を送信することができ、二相コミットにおける整合性確認や排他制御確認などを必要とせず遅延が発生しないようにすることができる。

【0175】

(実施形態13)

図32は、本実施形態における分散型データベースシステムにおけるデータの流れを説明するシーケンス図を例示する。

【0176】

計算機403で、キャッシュされているデータベースオブジェクトに対して更新が行なわれようとする、ステップS3201において、更新操作情報がトポロジ管理サーバ装置401へ送信される。トポロジ管理サーバ装置401は、トポロジ情報を参照して、更新がされようとしているデータベースオブジェクトを

管理しているデータベース管理装置を求め、そのデータベース管理装置へ更新操作情報を転送する（ステップS3202）。データベース管理装置402で、更新操作情報に基づいて、データベースオブジェクトが更新されると、実施形態12におけるように、データベースオブジェクトの更新操作命令が、トポロジ管理サーバ装置401へ送信され（ステップS3203）、計算機403へ転送され（ステップS3204）、計算機403で、データベースオブジェクトの更新が行なわれる。

【0177】

（実施形態13：更新操作が可能であるかどうかデータベース管理装置402で判断されてもよい）

なお、データベース管理装置402が更新操作情報を受信した際には、整合性のチェックや排他制御のチェックを行ない、整合性がとれなかったり排他制御に矛盾が生じたりする場合には、更新をしないようにしてもよい。このようにデータベース管理装置402が更新をしない場合には、エラーを示す情報が更新操作情報を送信した計算機403へ送信されてもよい。

【0178】

（実施形態13：データベース管理装置）

図33は、本実施形態におけるデータベース管理装置の機能ブロック図を例示する。本実施形態におけるデータベース管理装置は、実施形態2、実施形態9または実施形態11のデータベース管理装置が、更新操作情報受信部3301と、更新操作部3302と、更新操作命令送信部3303と、をさらに有している構成になっている。

【0179】

（実施形態13：データベース管理装置：更新操作情報受信部）

「更新操作情報受信部」3301は、更新操作情報を受信する。「更新操作情報」とは、データベースオブジェクトに対する更新操作に関する情報である。例えば、どのデータベースオブジェクトに対して、どのような更新の操作が行なわれようとしているか、あるいは、どのような更新の操作が行なわれるべきか、を表わす情報である。

【0180】

(実施形態13：データベース管理装置：更新操作部)

「更新操作部」3302は、更新操作情報受信部3301で受信された更新操作情報に基づいて自らが保持するデータベースオブジェクトに対して更新操作を行なう。「自らが保持するデータベースオブジェクト」とは、データベース管理装置402が保持するデータベースのデータベースオブジェクトである。

【0181】

(実施形態13：データベース管理装置：更新操作命令送信部)

「更新操作命令送信部」3303は、更新操作情報に基づいて、データベースオブジェクトをキャッシュしている計算機であるクライアント装置に向けて、キャッシュされているデータベースオブジェクトの更新操作の命令である更新操作命令を送信する。例えば、ブロードキャストによって更新操作命令を送信してもよい。もし、データベース管理装置が、データベースオブジェクトをキャッシュしている計算機を管理していれば、更新操作命令送信部から直接計算機へ更新操作命令が送信されてよい。

【0182】

(実施形態13：更新操作命令は、トポロジ管理サーバ装置を経由して計算機へ送られてもよい(実施形態9または11に基づく場合))

更新操作命令は、上で説明したようにデータベース管理装置から計算機へ直接送信されてもよい。あるいは、トポロジ管理サーバ装置へ送信されてもよい。トポロジ管理サーバ装置が管理されているトポロジ情報にデータベースオブジェクト識別子にそのデータベースオブジェクトがキャッシュされた計算機の識別子に対応付けている場合には、更新操作命令を受信したトポロジ管理サーバ装置は、更新されるべきデータベースオブジェクトをキャッシュしている計算機の識別子をトポロジ情報の参照により求め、求められた識別子を持つ計算機へ更新操作命令を送信する。

【0183】

(実施形態13：計算機)

図34は、本実施形態における計算機の機能ブロック図を例示する。本実施形

態における計算機は、実施形態 12 の計算機が更新操作情報送信部 3401 をさらに有している構成になっている。

【0184】

(実施形態 13：計算機：更新操作情報送信部)

「更新操作情報送信部」3401 は、更新操作情報を送信する。この送信のタイミングは、計算機 403 がキャッシュしているデータベースオブジェクトに対して更新操作を行なおうとするときである。このタイミングで、更新操作情報送信部 3401 は、どのデータベースオブジェクトに対して、どのような更新操作が行なわれようとしているかを検出して、更新操作情報を生成して送信する。なお、キャッシュしているデータベースオブジェクトが実際に更新されるのは、更新操作命令が更新操作命令受信部 3101 で受信され、データベースオブジェクト更新操作部 3102 により更新操作が行なわれるときである。

【0185】

(実施形態 13：更新操作情報はトポロジ管理サーバ装置を経由してデータベース管理装置へ送信されてもよい)

更新操作情報は、データベース管理装置へ直接送信されてもよい。あるいは、一度トポロジ管理サーバ装置へ送られてもよく、トポロジ管理サーバ装置はトポロジ情報を参照して、更新されようとしているデータベースオブジェクトを管理しているデータベース管理装置へ更新操作情報を送信する。また、トポロジ管理サーバ装置が他のトポロジ管理サーバ装置とトポロジ情報を交換している場合には、もし、更新操作情報に係るデータベースオブジェクトが他のドメインのデータベース管理装置が管理するデータベースオブジェクトであれば、他のドメインのトポロジ管理サーバ装置へ更新操作情報を転送する。

【0186】

(実施形態 13：主な効果)

本実施形態によれば、クライアント装置である計算機でのデータベースオブジェクトに対する更新操作が、データベース管理装置で管理されているデータベースオブジェクトや、他の計算機にキャッシュされているデータベースオブジェクトに反映されることになる。これにより、例えば、計算機で動作するプログラム

の結果を他の計算機で動作するプログラムへ伝えることが可能となる。

【0187】

また、本実施形態においては、計算機がキャッシュしているデータベースオブジェクトを更新することが可能である。更新は、データベース管理装置での更新が正常に完了した場合にのみ、そのデータベースオブジェクトをキャッシュしている計算機だけにその更新内容を送信することが可能である。これにより、従来のレプリケーション技術を使用している場合のように、キャッシュしているデータを変更できないという不具合や、キャッシュしているデータを変更できないことにより、同期対象を絞り込めないために遅延を生ずるという不具合を避けることができる。

【0188】

(実施形態14)

本発明の実施形態14においては、データベースを停止させることなく、バックアップが行なわれ、障害が発生した場合に復旧が可能となる分散型データベースシステムが提供される。このために、データベース管理装置で生成される更新ジャーナルをネットワークに接続された物理的に異なるサーバ装置に出力する。

【0189】

(実施形態14：構成)

図35は、本実施形態の分散型データベースシステムの機能ブロック図を例示する。分散型データベースシステムは、実施形態1の分散型データベースシステムがジャーナル管理サーバ装置3501を備えた構成を採っている。

【0190】

図36は、ジャーナル管理サーバ装置3501の機能ブロック図を例示する。ジャーナル管理サーバ装置3501は、ジャーナル受信部3601と、ジャーナル蓄積部3602と、ジャーナルリプレイ部3603とスナップショット保持部3604と、リカバリ部3605とを有する。

【0191】

図37は、本実施形態におけるデータベース管理装置の機能ブロック図を例示しており、実施形態1のデータベース管理システムがジャーナル送信部3701

を有した構成になっている。

【0192】

(実施形態 14：ジャーナル管理サーバ装置：ジャーナル受信部)

「ジャーナル受信部」3601は、ジャーナルを受信する。「ジャーナル」とは、データベース管理装置で管理されるデータベースオブジェクトに対する更新を表わす情報である。すなわち、データベース管理装置でデータベースオブジェクトに対して、どのような更新の操作が行なわれたかを表わす情報である。ジャーナルは、個々の更新の操作ごとに生成されてもよい。あるいは、ジャーナルは、トランザクションがコミットされるなどのタイミングで、一つまたは複数の更新の操作が行なわれるごとに生成されてもよい。

【0193】

(実施形態 14：ジャーナル管理サーバ装置：ジャーナル蓄積部)

「ジャーナル蓄積部」3602は、ジャーナル受信部3601で受信されたジャーナルを蓄積する。例えば、メモリや磁気ディスク、光ディスクなどに蓄積する部である。あるいは、電源などの信頼性が高ければ主メモリにジャーナルを蓄積するようになっていてもよい。

【0194】

(実施形態 14：ジャーナル管理サーバ装置：ジャーナルリプレイ部)

「ジャーナルリプレイ部」3603は、ジャーナル蓄積部3602で蓄積したジャーナルをリプレイする。「リプレイ」とは、ジャーナルが表わすデータベースオブジェクトに対する更新の操作を、ジャーナル管理サーバ装置3501で、実行することを意味する。ジャーナルのリプレイの実行は、スナップショット保持部3604に保持されているスナップショットに対して実行される。

【0195】

このリプレイは、ジャーナル蓄積部3602にジャーナルが蓄積される都度、行なわれてもよい。あるいは、ジャーナル蓄積部3602に所定の量以上のジャーナルが蓄積された時に、リプレイが行なわれるようになっていてもよい。あるいは、所定の時間ごとにリプレイが行なわれるようになっていてもよい。

【0196】

「スナップショット保持部」3604は、ジャーナルリプレイ部3603でリプレイされたジャーナルに基づいて生成されたスナップショットを保持する。

【0197】

ジャーナルがリプレイされることにより、データベース管理装置の管理されるデータベースがジャーナル管理サーバ装置において再現される。「スナップショット」とは、このように再現されたデータベースのある時点における複製である。このような複製は、例えば、メモリや磁気ディスク装置、光ディスク装置などにより記憶され保持される。また、スナップショットが保持される都度、ジャーナル蓄積部3602からリプレイされたジャーナルを削除するようにしてもよい。

【0198】

なお、スナップショットは、複数保持されていてもよい。例えば、(1)特定のジャーナルがリプレイされる前のスナップショット、(2)特定のジャーナルがリプレイされた後のスナップショット、など、2つ以上のスナップショットが保持されてもよい。

【0199】

「リカバリ部」3605は、ドメインが障害を蒙ったときに前記スナップショットから障害ドメインのリカバリのための処理を行なう機能を有する部である。「ドメインが障害を蒙ったとき」には、例えば、分散型データベースシステムのデータベース管理装置に障害が発生したとき、がある。「障害ドメイン」とは、障害を蒙ったドメインである。「リカバリのための処理」とは、障害から復旧するための処理である。例えば、スナップショット保持部に保持されたスナップショットをデータベース管理装置へ送信し、スナップショット保持部にスナップショットが保持された後にジャーナル蓄積部に蓄積されたジャーナルをデータベース管理装置にてリプレイする。あるいは、スナップショット保持部に保持されたスナップショットに対して、スナップショット保持部にスナップショットが保持された後にジャーナル蓄積部に蓄積されたジャーナルをリプレイして得られるスナップショットをデータベース管理装置へ送信する。あるいは、新しくデータベース管理装置を用意して、そのデータベース管理装置へスナップショットを送信

するようにしてもよい。

【0200】

(実施形態14：データベース管理装置：ジャーナル送信部)

「ジャーナル送信部」3701は、ジャーナルを送信する。すなわち、データベース管理装置402でのデータベースオブジェクトに対してどのような更新の操作が行なわれたかを表わす情報を送信する。この送信は、データベースオブジェクトに対して更新の操作が行なわれる都度行なわれてもよい。あるいは、トランザクションがコミットされるなど、一定のイベントが発生する都度行なわれてもよい。

【0201】

(実施形態14：主な効果)

本発明においては、データベースオブジェクトが計算機にキャッシュされているため、データベースを停止させることが困難であるので、本実施形態により、データベースを停止させることなくデータベースのバックアップを採ることができる。また、スナップショットの移動により障害からのリカバリが行なわれるので、復旧を短時間で終わらせることができる。

【0202】

また、データベース管理装置などのハードウェアの故障やソフトウェアのハングアップなどによる再起動に伴うメインメモリ上のデータの消失に対処することができる。リカバリの実行は限られたドメイン内で完結するため、巨大データベースのリカバリ処理も分散したオブジェクトで完結するため運用上の負担が減る。

【0203】

(実施例1)

以下に、本発明の実施例を挙げる。

【0204】

企業内に設置されたワークステーションあるいはパーソナルコンピュータがLANに接続されている。従業員の机上のパーソナルコンピュータは、就業時間中に使用されるが、夜間や休日は使用されていない。それらのパーソナルコンピュ

ータは高性能であるにもかかわらず、稼動するソフトウェアは、ワードプロセッサ、表計算、プレゼンテーション作成ツール、メーラ、ブラウザなどの計算資源をそれほど要求しないものであり、CPU能力、メインメモリ容量、磁気ディスク容量に余剰が生じている。

【0205】

一方、月次の請求入金処理が月末に集中するので、パーソナルコンピュータの余剰能力を利用するために、本発明の分散型データベースの計算機として、これらのパーソナルコンピュータを使用する。この場合、パーソナルコンピュータの負荷が一定以下になるものに、請求入金処理のためのデータベースオブジェクトをキャッシュさせて、データベースオブジェクトを参照して請求入金処理を行なうプログラムを動作させる。これにより、ワークステーションなどを増強することなしに月末処理を行なうことができる。

【0206】

(実施例2)

また、別の実施例を以下に挙げる。

【0207】

マンションなどの集合住宅にブロードバンドインターネットサービスを提供する会社が、その集合住宅の全戸にサービスを利用してもらうために、利用者からサービス利用料を徴収しないことにした。その代わりに、全戸に低消費電力の高性能パーソナルコンピュータを配布し、常時電源を入れた状態にしてもらうことを条件とした。もちろん、ブロードバンドインターネットにも常時接続されることが条件となっている。

【0208】

配布した低消費電力の高性能パーソナルコンピュータを、本発明の分散データベースシステムの計算機とする。この高性能パーソナルコンピュータは、回転軸などの摺動機械要素を含む磁気ディスクを含まず、機械的な故障ができるだけ発生しないようにしたものであってもよい。また、停電などに備えて、無停電装置に接続されていてもよい。ブロードバンドインターネットサービスを提供する会社は、コンピュータ資源を必要としている会社と契約を行ない、全戸に配布され

た低消費電力の高性能パーソナルコンピュータの余剰コンピュータ資源をまとめて提供することにする。この余剰コンピュータ資源の使用料を、ブロードバンドインターネットサービスを提供する会社が、契約した会社から徴収する。

【0209】

また、データベースオブジェクトを用いたグループウェアのソフトウェアを各戸のパーソナルコンピュータで動作させることにより、集合住宅内のグループウェア環境が実現され、また、地域情報ネットワークが実現される。

【0210】

このような地域情報ネットワークは、集合住宅をドメインとするトポロジ管理サーバ装置間でトポロジ情報が交換されることにより、成長し、マーケティングリソースとしてその価値を大きくして行く。

【0211】

【発明の効果】

以上のように、本発明によれば、データベースオブジェクトを複数の計算機に分散することができる。また、計算機のCPU資源、メモリ資源を有効に用いて、計算を分散して行なうことが可能となる。また、データベースを停止させることなく、データベースのバックアップが行なえる。

【図面の簡単な説明】

【図1】

本発明の概念図

【図2】

計算機からキャッシュ要求が送信され、それがトポロジ管理サーバ装置に転送されて、データベース管理装置からデータベースオブジェクトの転送がされる過程を示すシーケンス図

【図3】

計算機が、キャッシュしているデータベースオブジェクトにロックを行なう際の過程に関するシーケンス図

【図4】

実施形態1における分散型データベースシステムの機能ブロック図

【図 5】

実施形態 1 におけるトポロジ管理サーバ装置 401 の機能ブロック図

【図 6】

キャッシュ要求の模式図

【図 7】

実施形態 1 における計算機と、トポロジ管理サーバ装置と、データベース管理装置と、の間の処理の流れを説明するシーケンス図

【図 8】

実施形態 1 におけるトポロジ管理サーバ装置の処理の流れを説明するフローチャート

【図 9】

実施形態 1 における分散型データベースシステムの別の機能ブロック図

【図 10】

計算機がどのデータベースオブジェクトに対するキャッシュ要求を送信したかを管理するために用いられる表の一例図

【図 11】

実施形態 1 における計算機の機能ブロック図

【図 12】

実施形態 1 における算機の処理の流れを説明するフローチャート

【図 13】

実施形態 1 におけるデータベース管理装置の機能ブロック図

【図 14】

実施形態 1 データベース管理装置の処理の流れを説明するフローチャート

【図 15】

実施形態 2 における分散型データベースシステムの計算機の機能ブロック図

【図 16】

実施形態 3 における分散型データベースシステムの計算機の機能ブロック図

【図 17】

実施形態 5 におけるキャッシュ要求の模式図

【図 18】

実施形態 6 におけるキャッシュ要求の模式図

【図 19】

実施形態 7 におけるキャッシュ要求の模式図

【図 20】

実施形態 8 におけるキャッシュ要求の模式図

【図 21】

実施形態 9 におけるトポロジ情報の模式図

【図 22】

図 21 に模式的に例示されたトポロジ情報を表現する表の一例図

【図 23】

データベースオブジェクト識別子とロック情報を対応付けている表の一例図

【図 24】

実施形態 9 におけるトポロジ管理サーバ装置の機能ブロック図

【図 25】

計算機がキャッシュ済情報送信部を有している場合のデータの流を例示するシーケンス図

【図 26】

実施形態 10 におけるトポロジ管理サーバ装置の機能ブロック図

【図 27】

実施形態 11 の概要を示す図

【図 28】

実施形態 11 におけるトポロジ管理サーバ装置の機能ブロック図

【図 29】

実施形態 12 におけるデータの流を説明するシーケンス図

【図 30】

実施形態 12 におけるデータベース管理装置の機能ブロック図

【図 31】

実施形態 12 における計算機の機能ブロック図

【図 3 2】

実施形態 1 3 における分散型データベースシステムにおけるデータの流れを説明するシーケンス図

【図 3 3】

実施形態 1 3 におけるデータベース管理装置の機能ブロック図

【図 3 4】

実施形態 1 3 における計算機の機能ブロック図

【図 3 5】

実施形態 1 4 における分散型データベースシステムの機能ブロック図

【図 3 6】

実施形態 1 4 におけるジャーナル管理サーバ装置の機能ブロック図

【図 3 7】

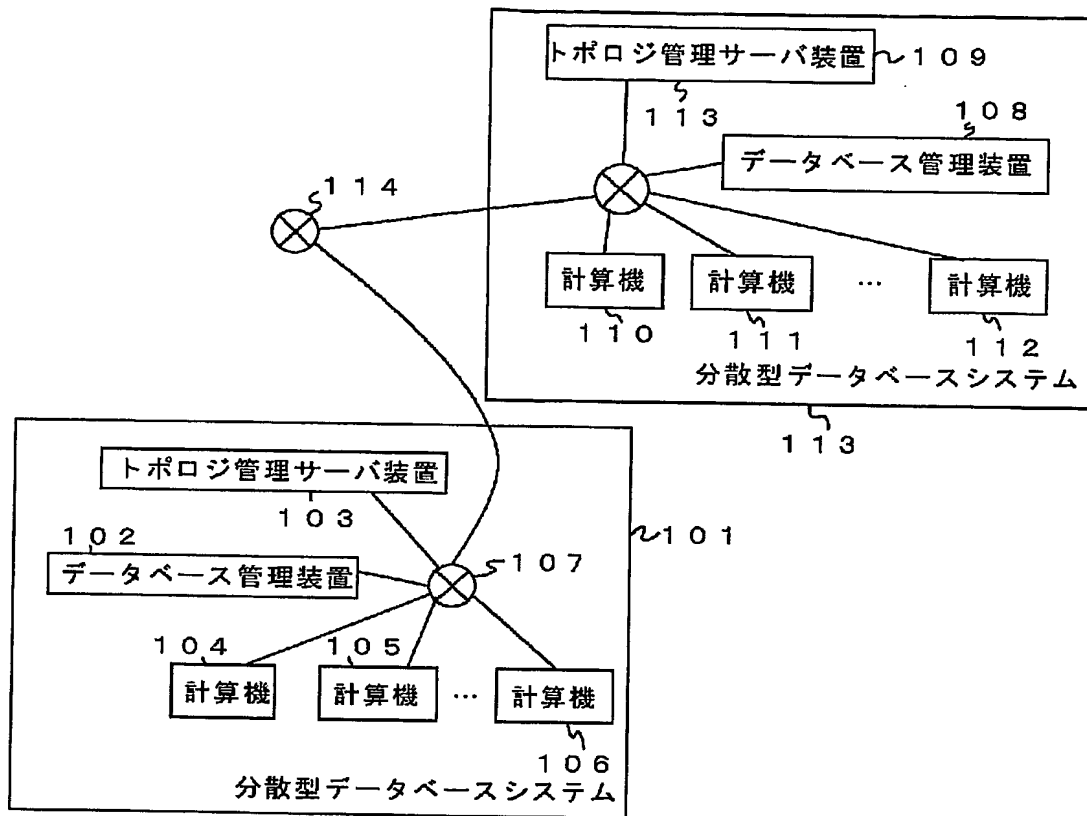
実施形態 1 4 におけるデータベース管理装置の機能ブロック図

【符号の説明】

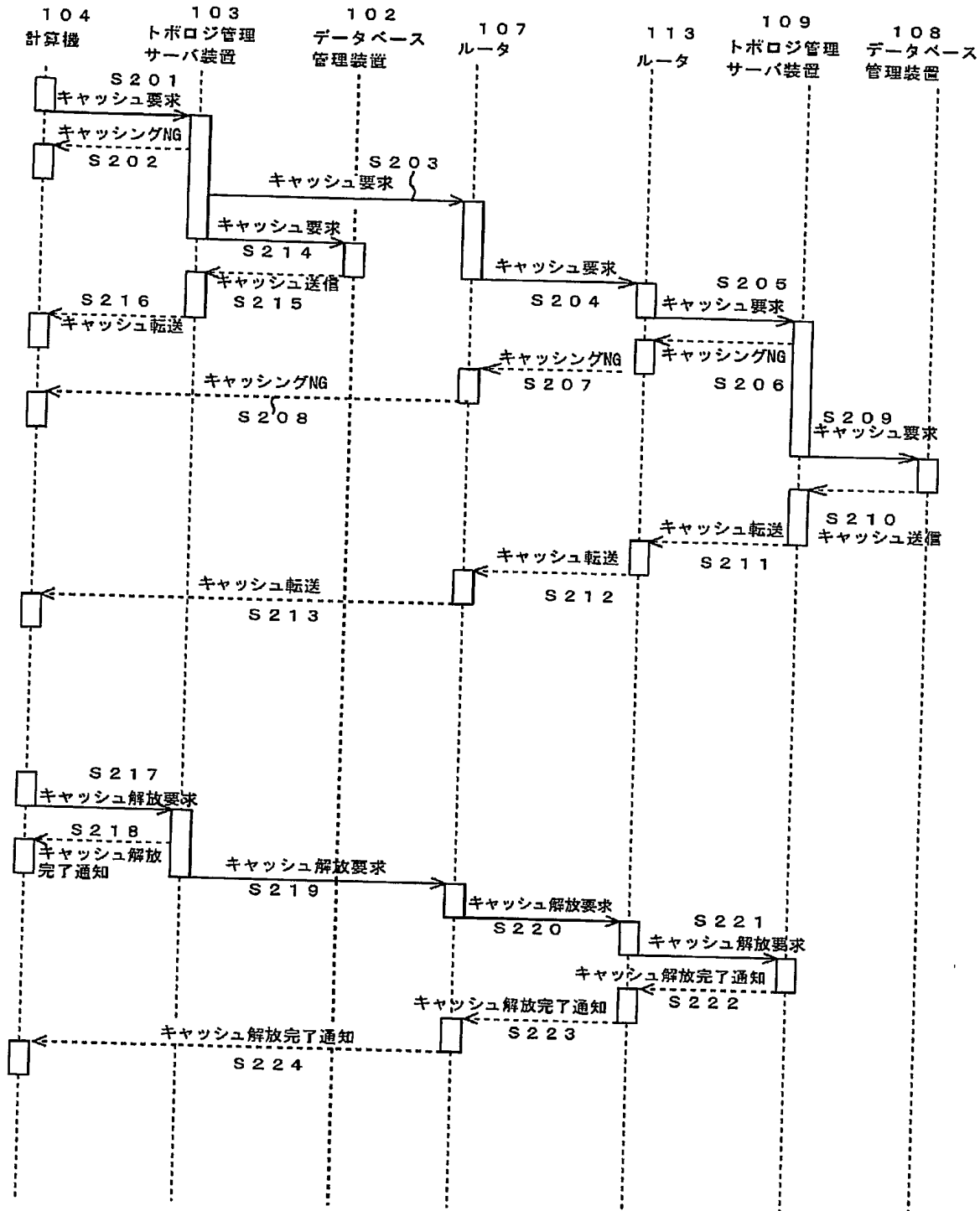
- 1 0 1 分散型データベースシステム
- 1 0 2 データベース管理装置
- 1 0 3 トポロジ管理サーバ装置
- 1 0 4 計算機
- 1 0 5 計算機
- 1 0 6 計算機
- 1 0 7 ルータ
- 1 0 8 データベース管理装置
- 1 0 9 トポロジ管理サーバ装置
- 1 1 0 計算機
- 1 1 1 計算機
- 1 1 2 計算機
- 1 1 4 通信網

【書類名】 図面

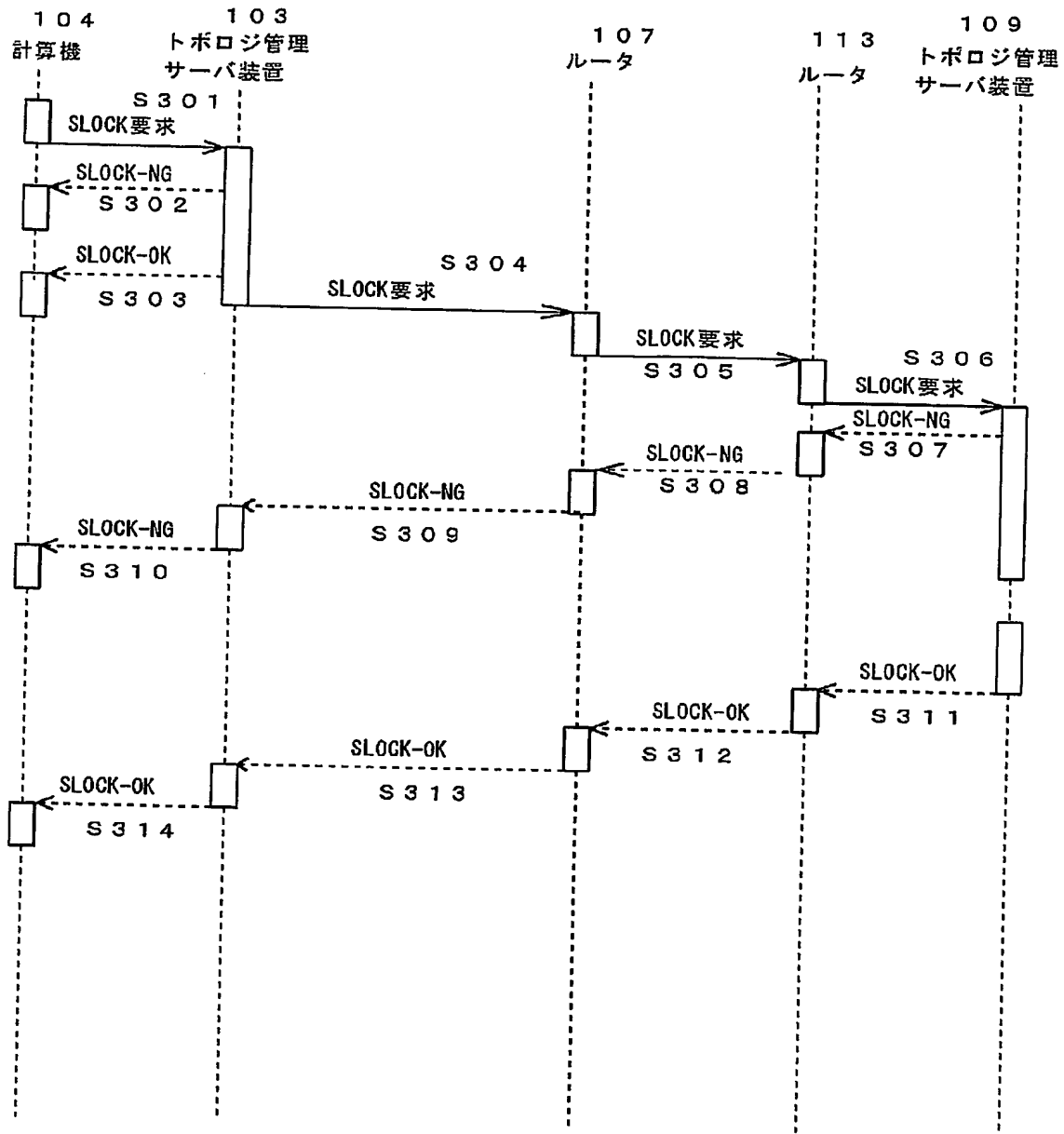
【図 1】



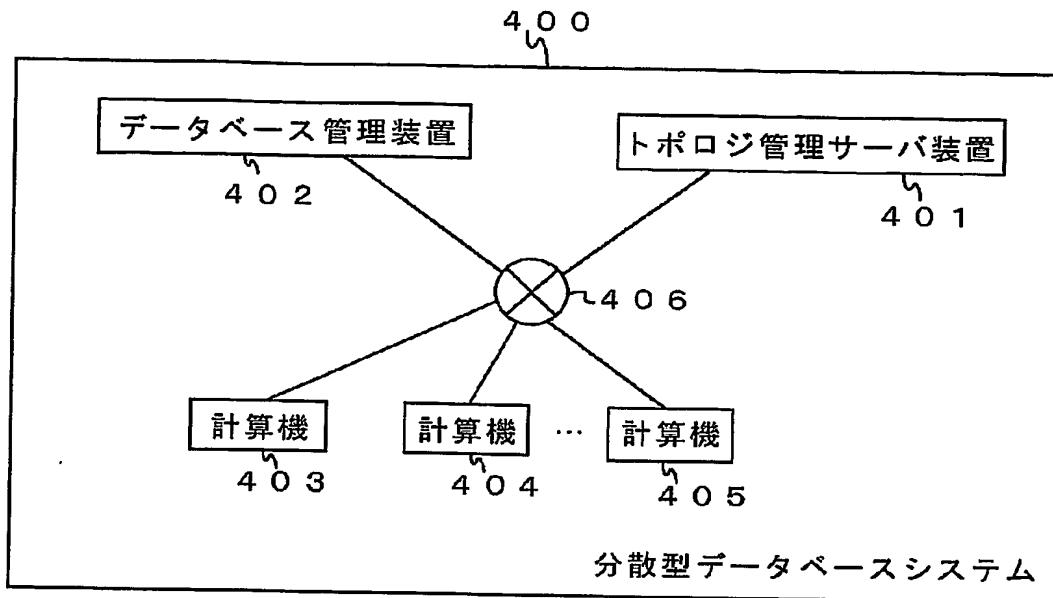
【図 2】



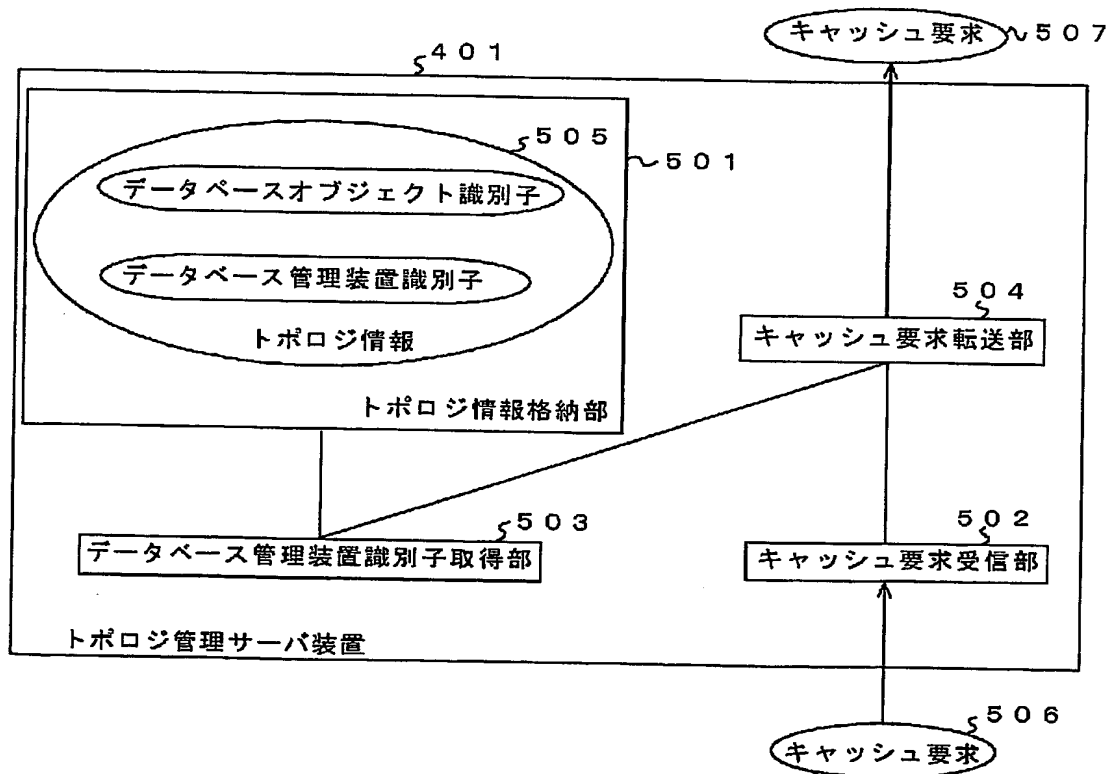
【図 3】



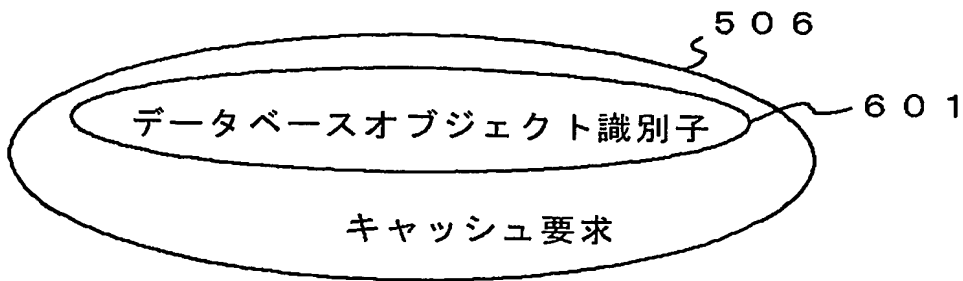
【図 4】



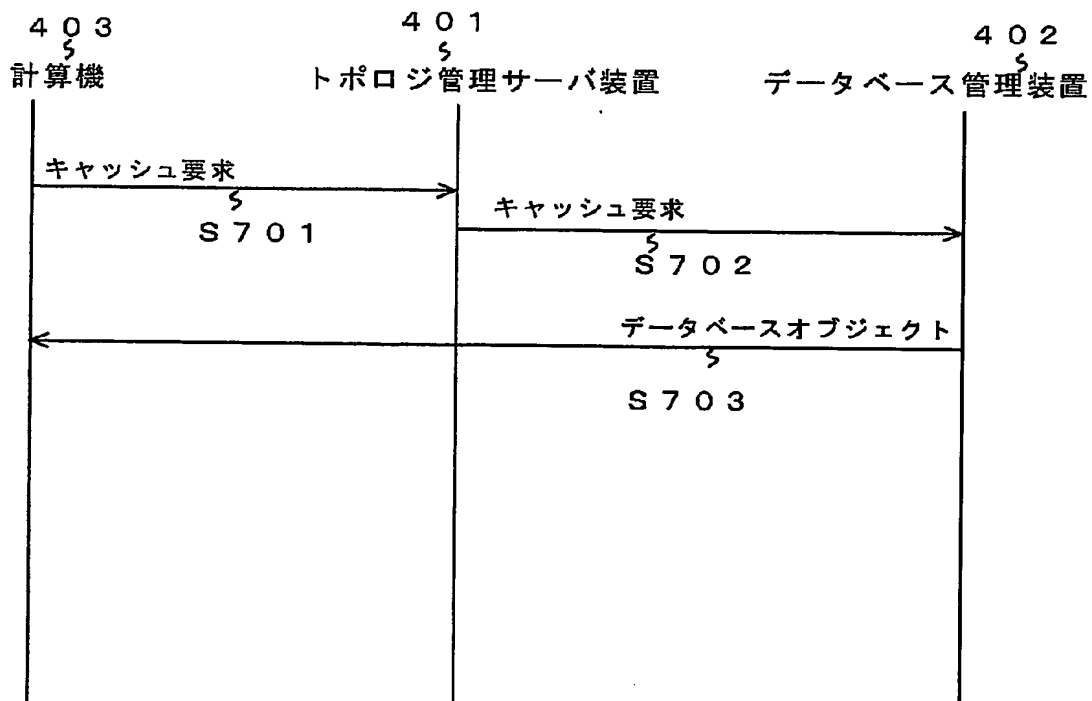
【図 5】



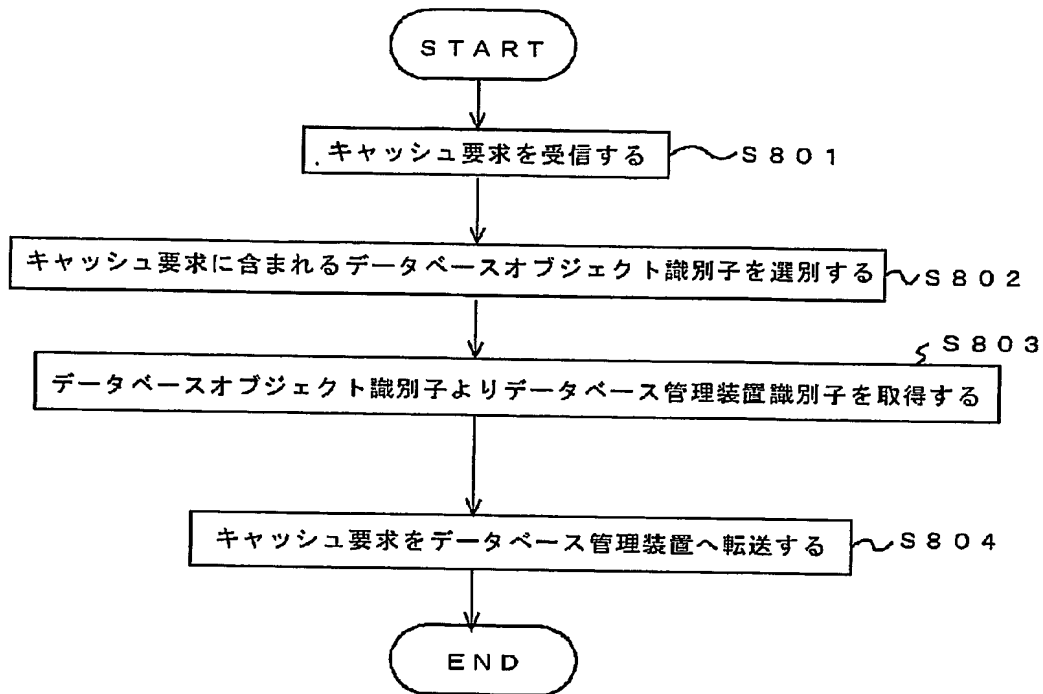
【図 6】



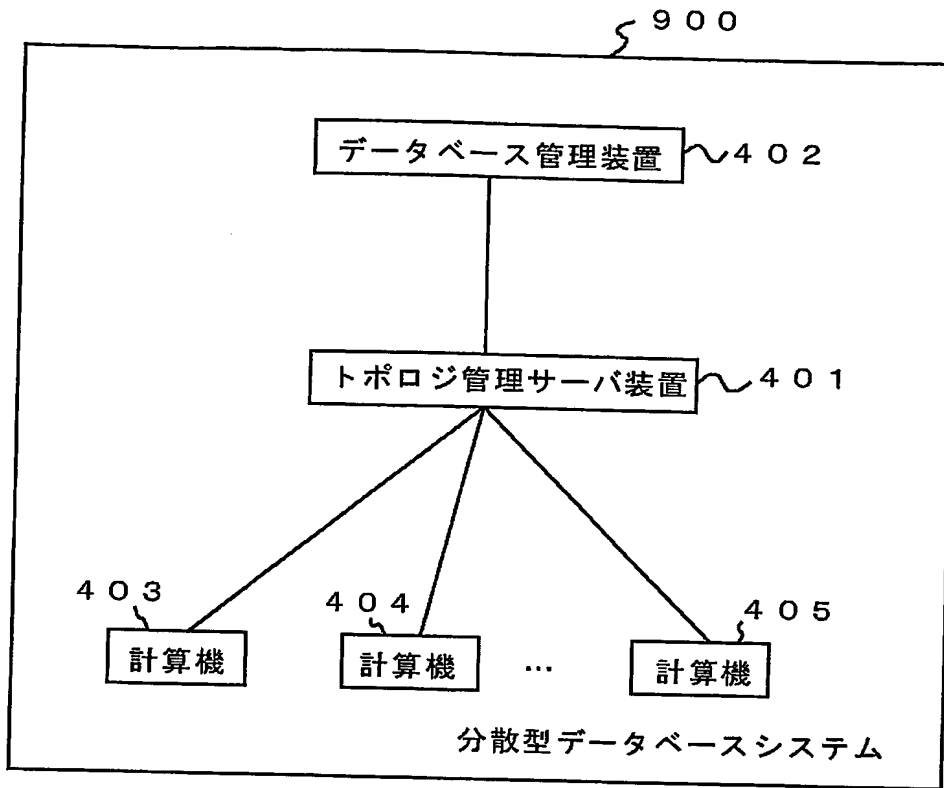
【図 7】



【図 8】



【図 9】

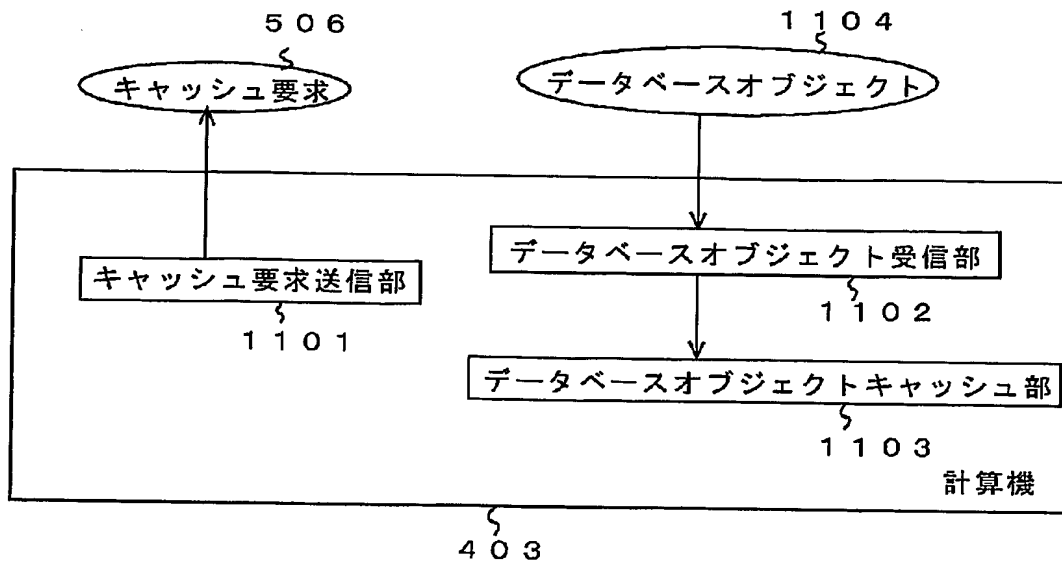


【図 10】

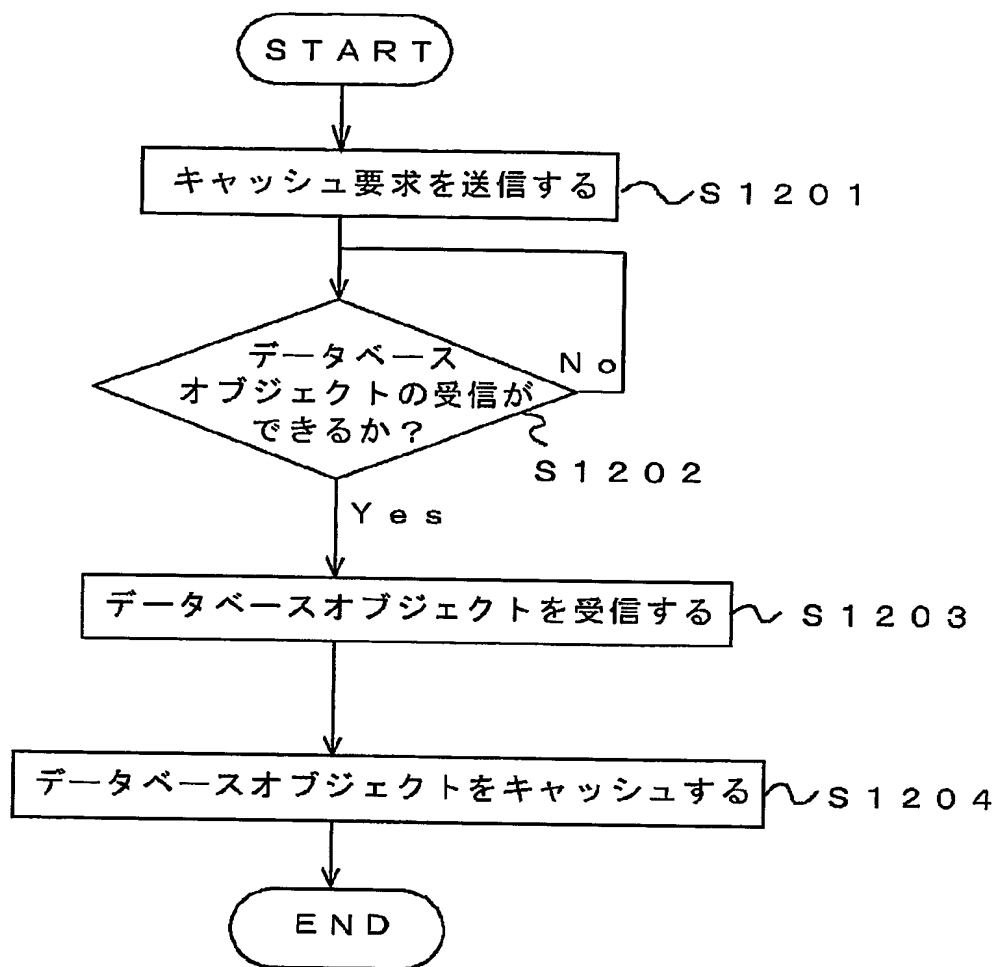
§ 1001

キャッシュ要求送信元	データベースオブジェクト識別子
133.139.210.3	DB34-T284729
133.139.213.210	DB473-T4055-t9029
⋮	⋮

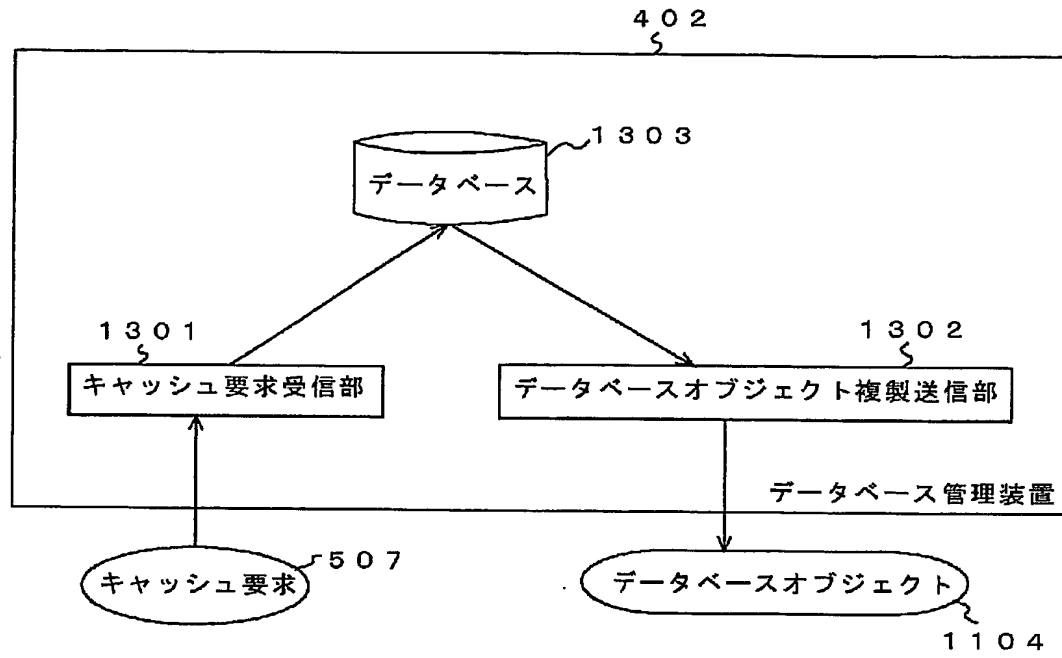
【図 11】



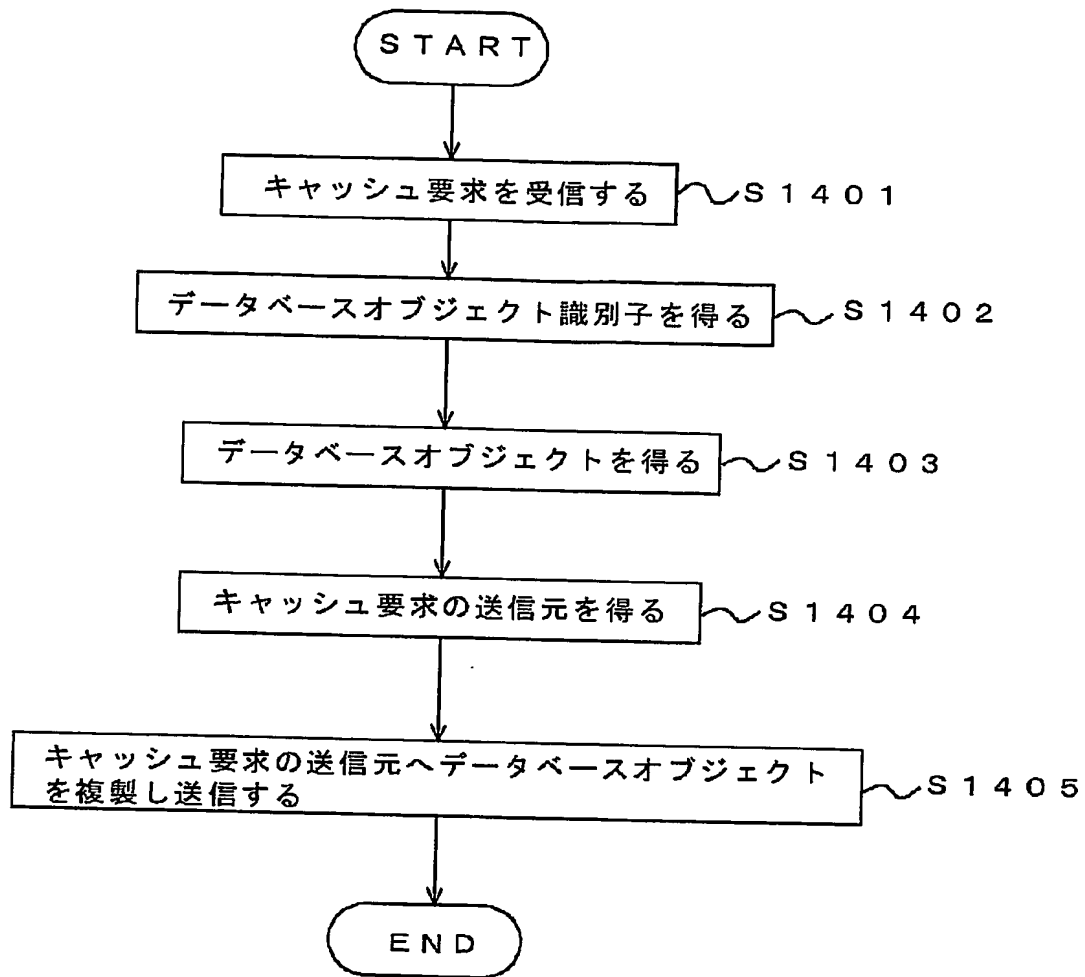
【図 12】



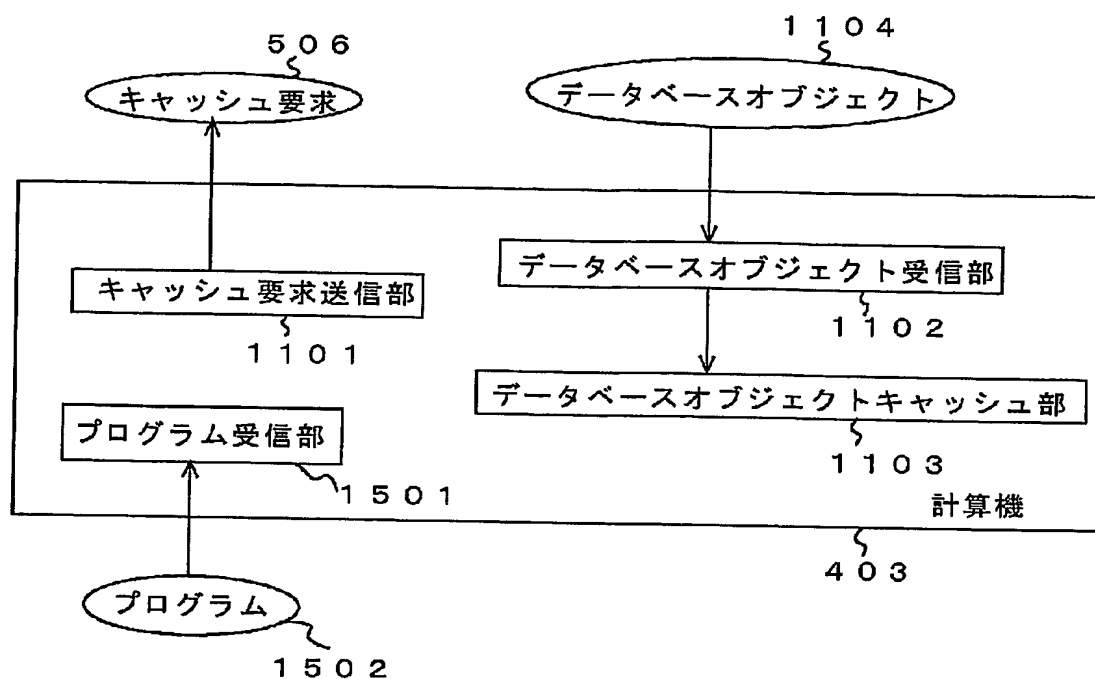
【図 13】



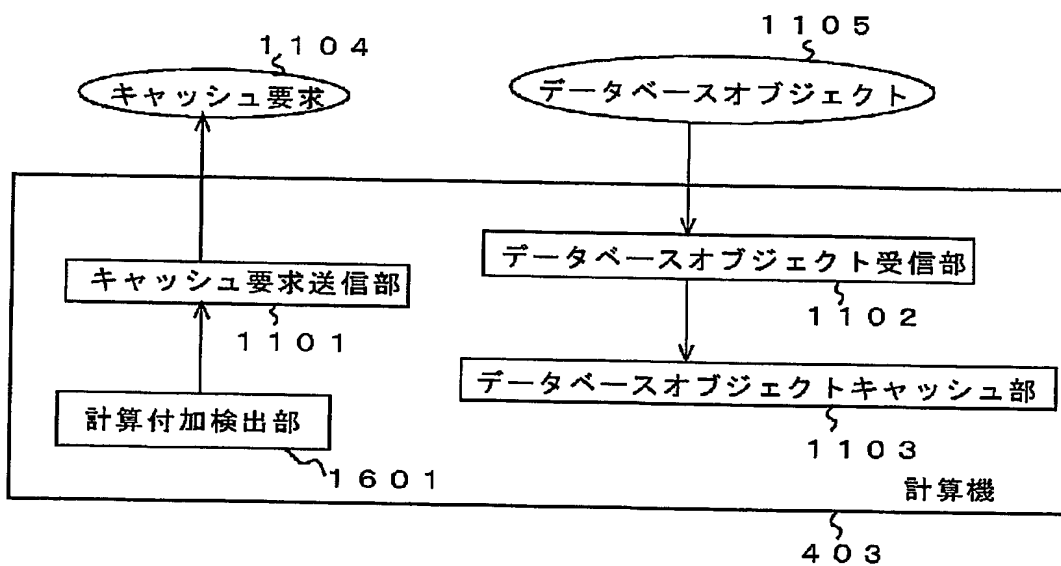
【図 14】



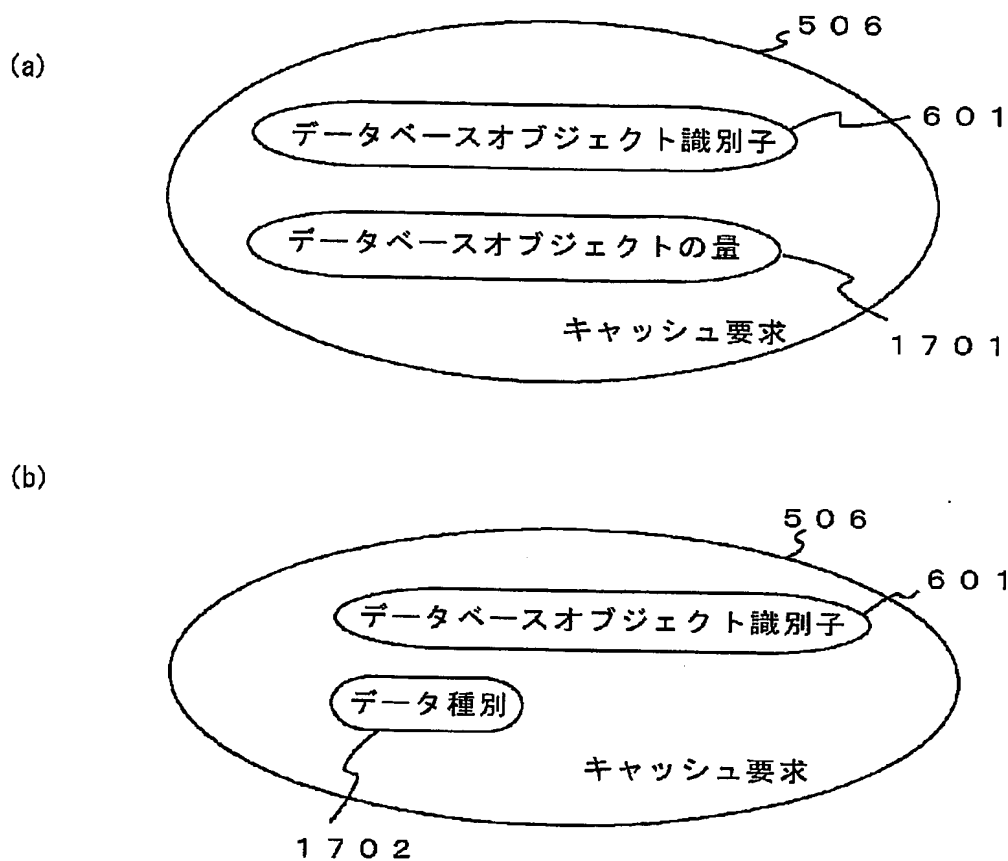
【図 15】



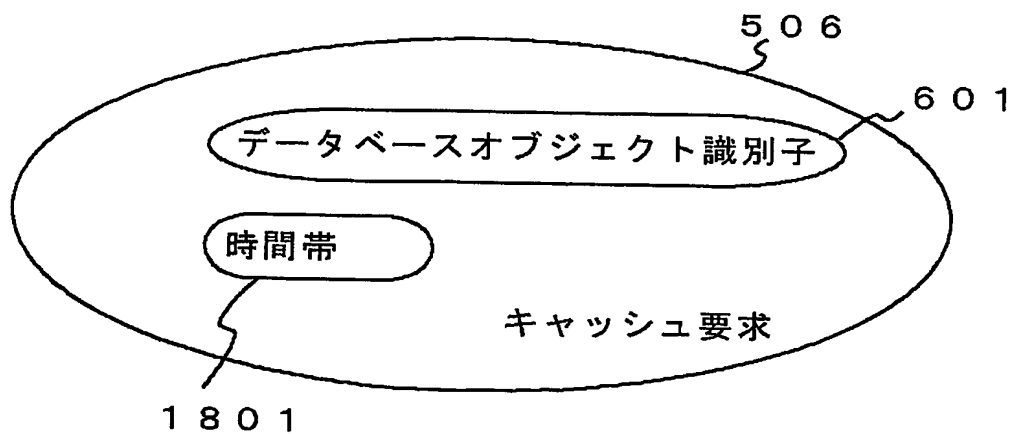
【図 16】



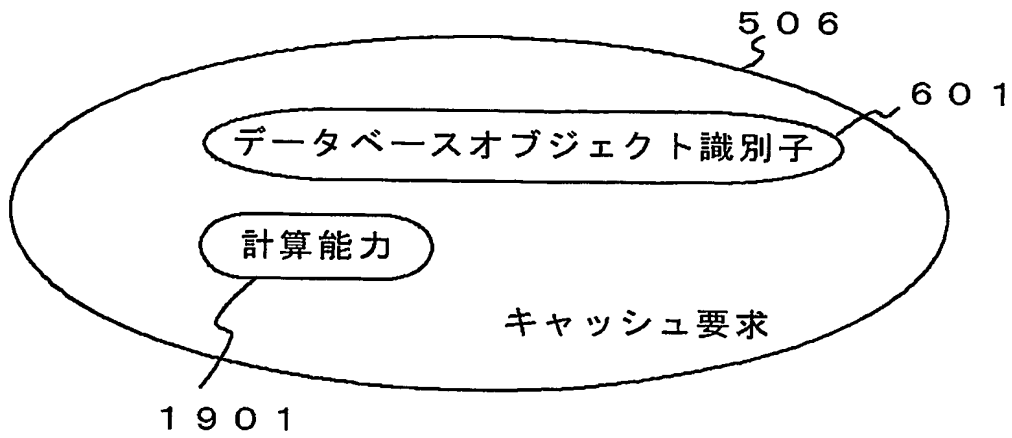
【図 17】



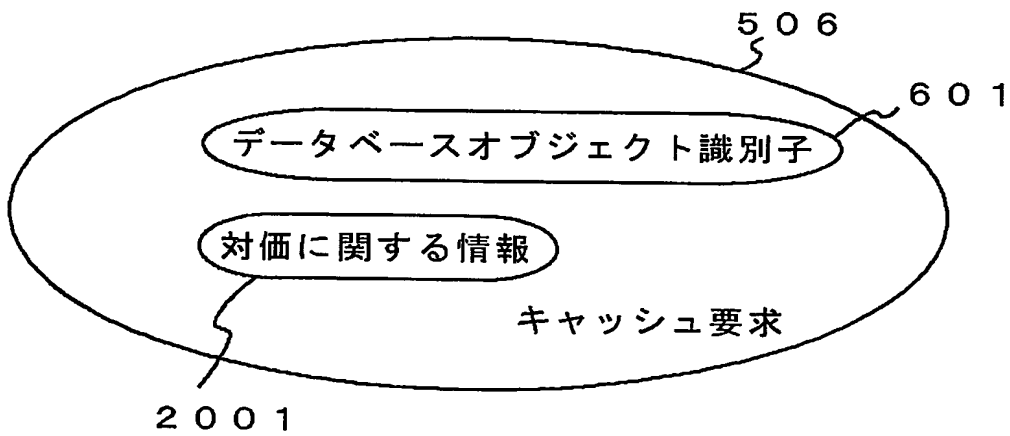
【図 18】



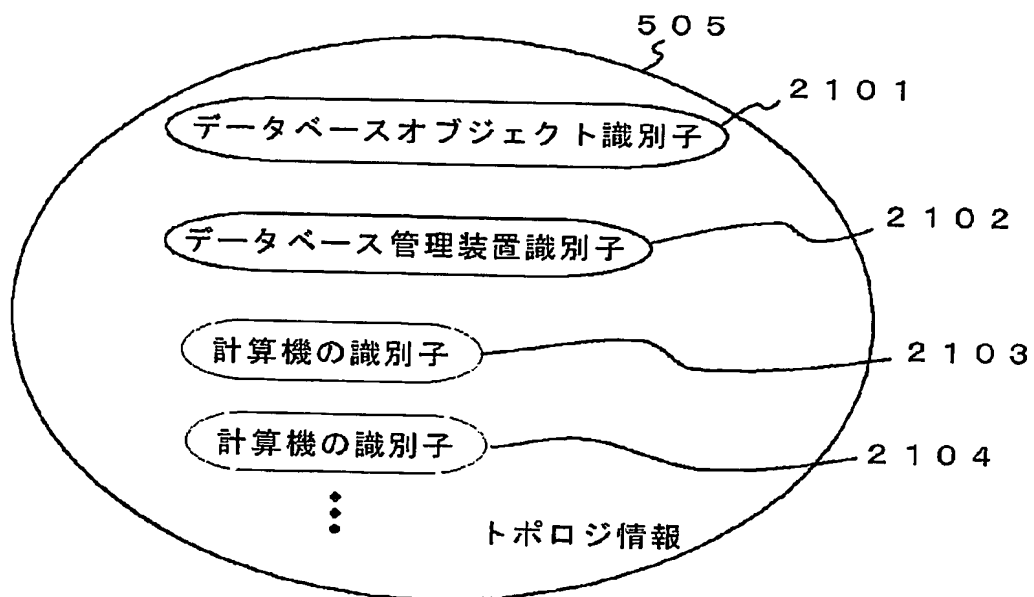
【図 19】



【図 20】



【図 21】



【図 22】

52201

データベースオブジェクト識別子	データベース管理装置識別子
DB34-T284729	10, 5, 234, 109
⋮	⋮

52202

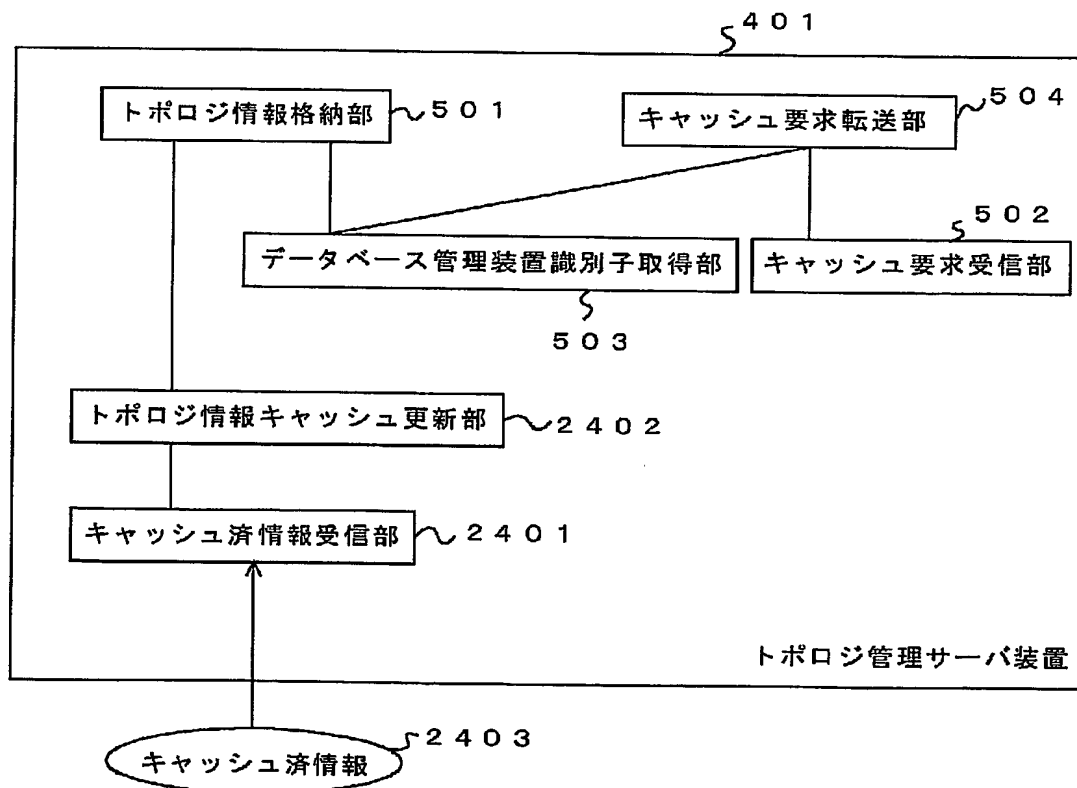
データベースオブジェクト識別子	計算機の識別子
DB34-T284729	133. 139. 210. 3
DB34-T284729	133. 139. 213. 210
⋮	⋮

【図 23】

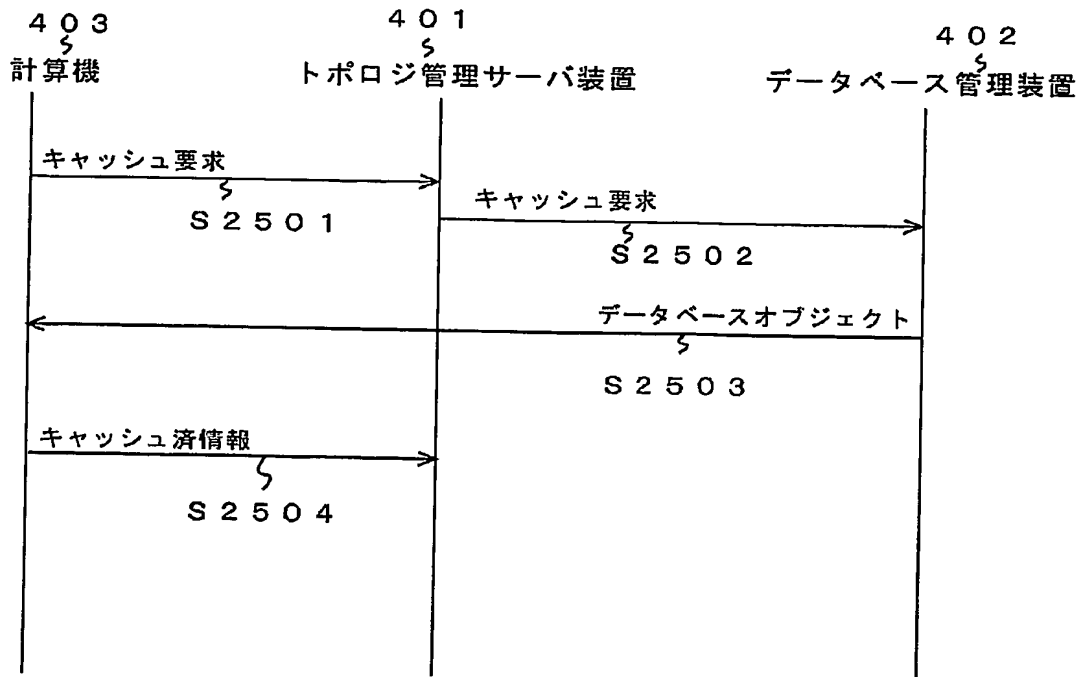
52301

データベースオブジェクト識別子	ロック情報
DB34-T284729	...
⋮	⋮

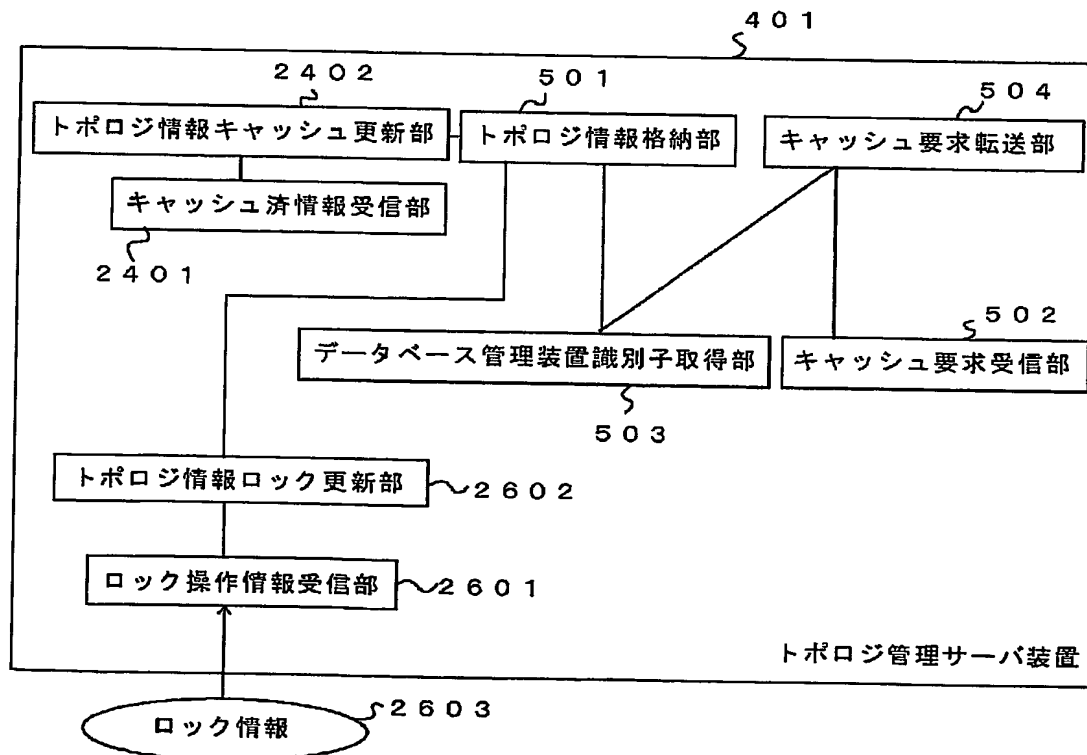
【図 24】



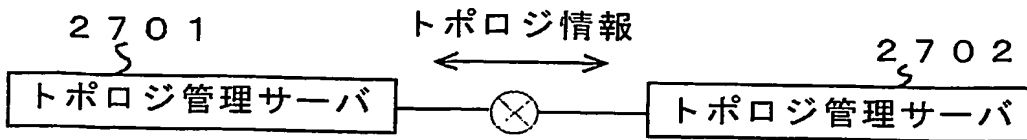
【図 25】



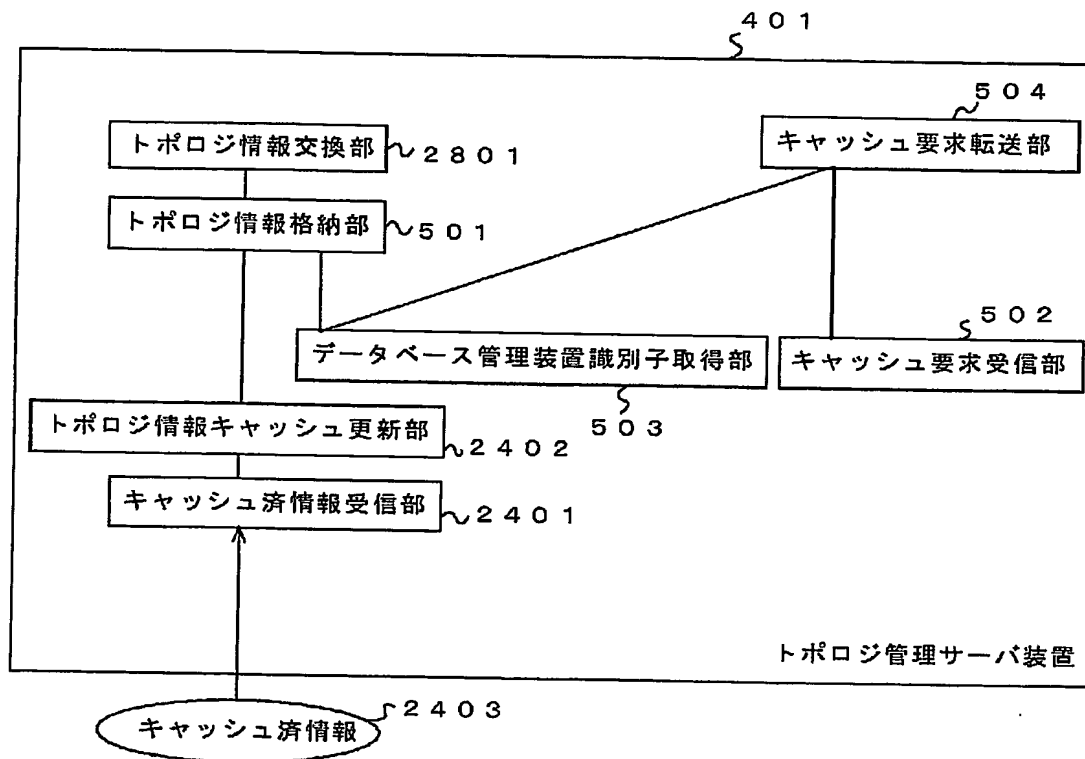
【図 26】



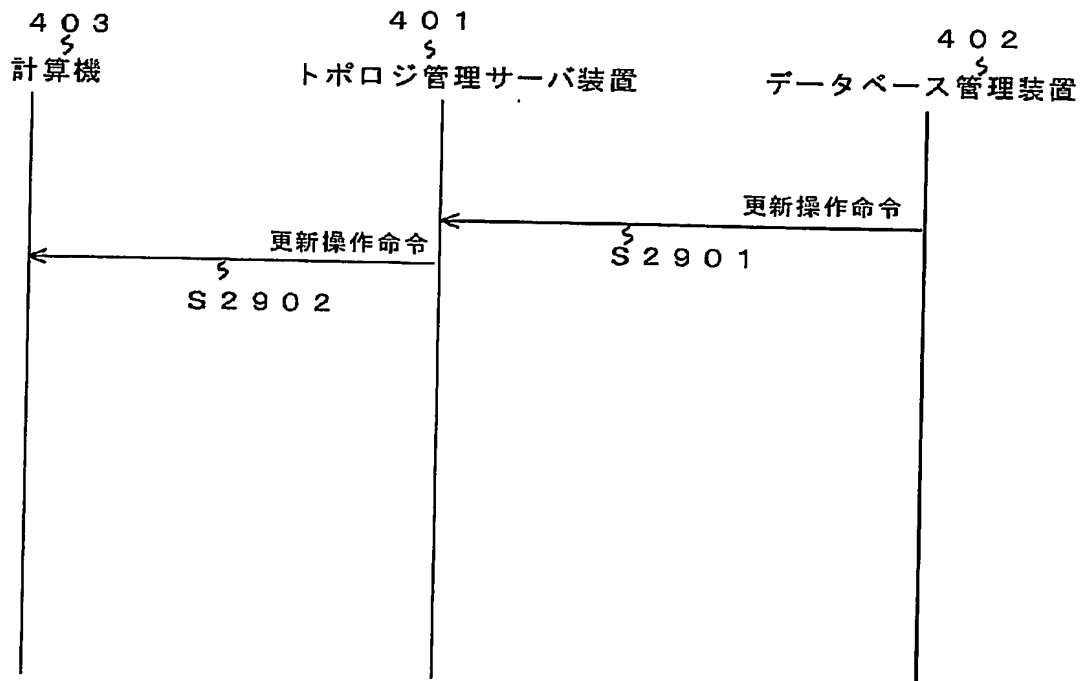
【図 27】



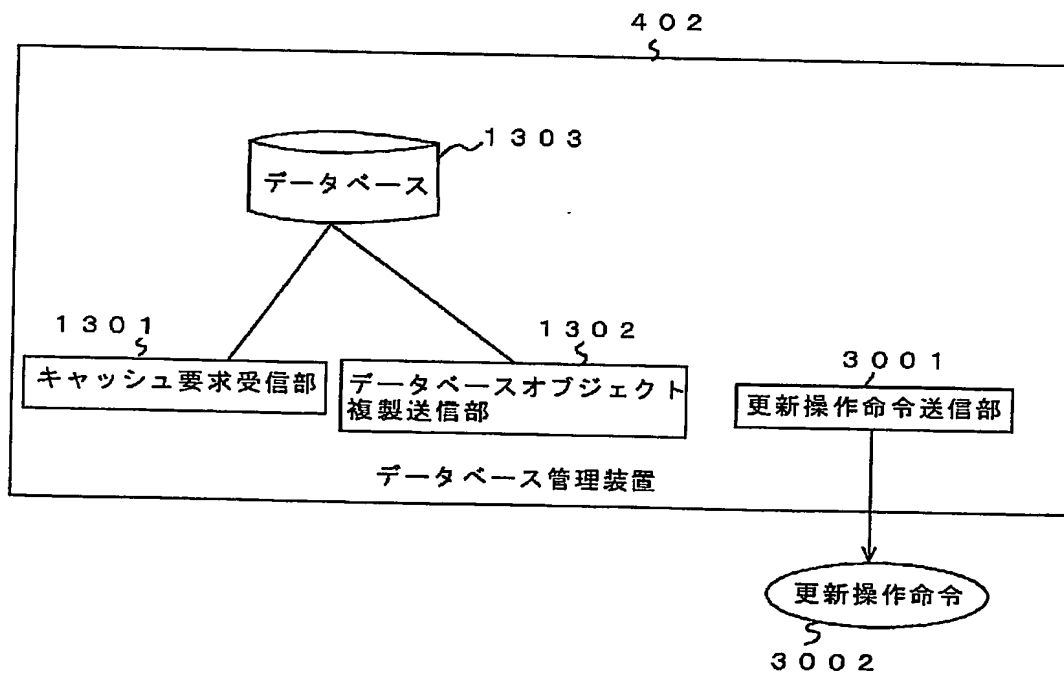
【図 28】



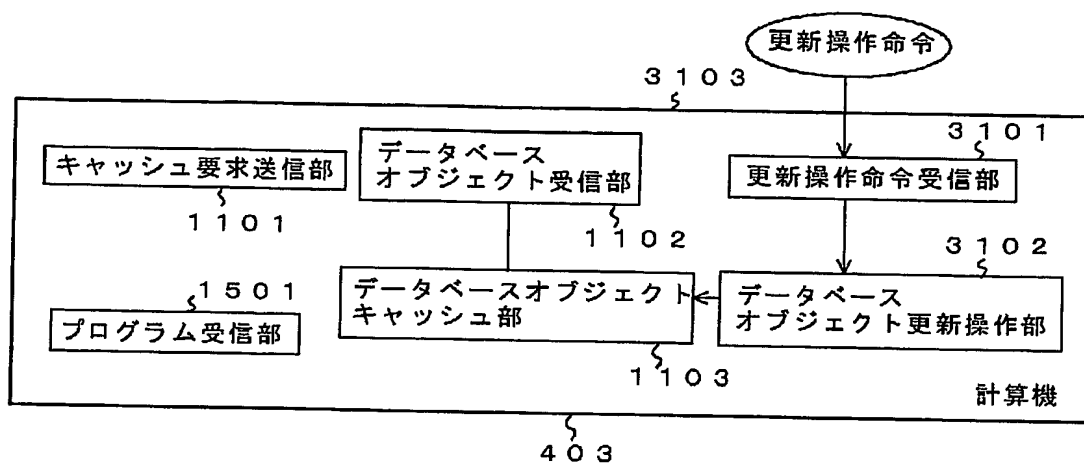
【図 29】



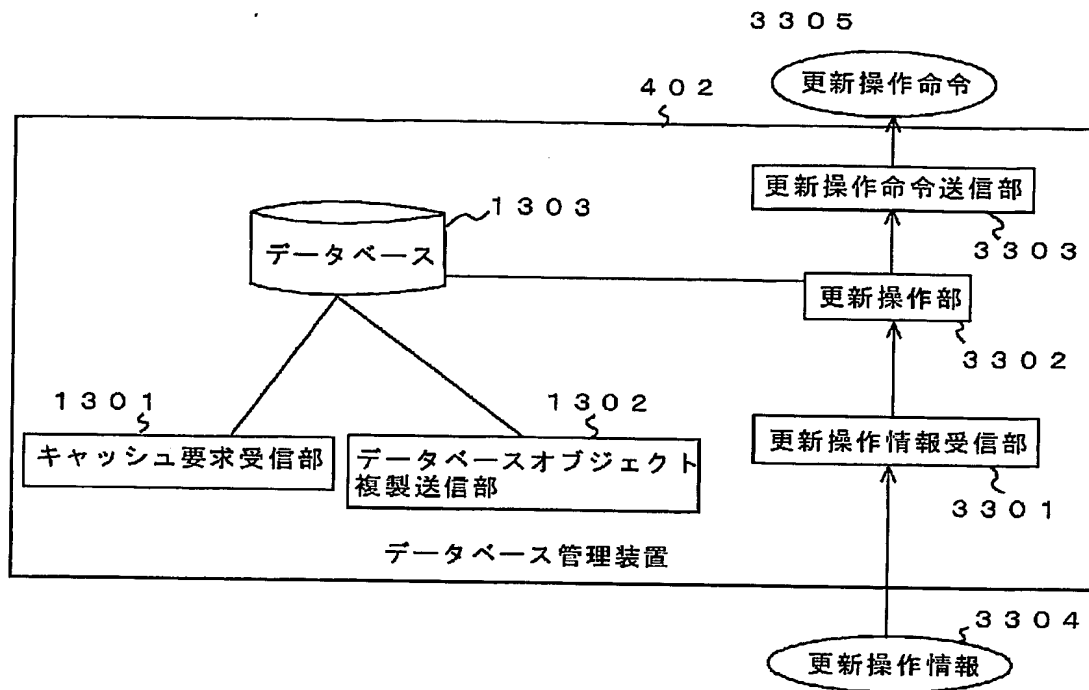
【図 30】



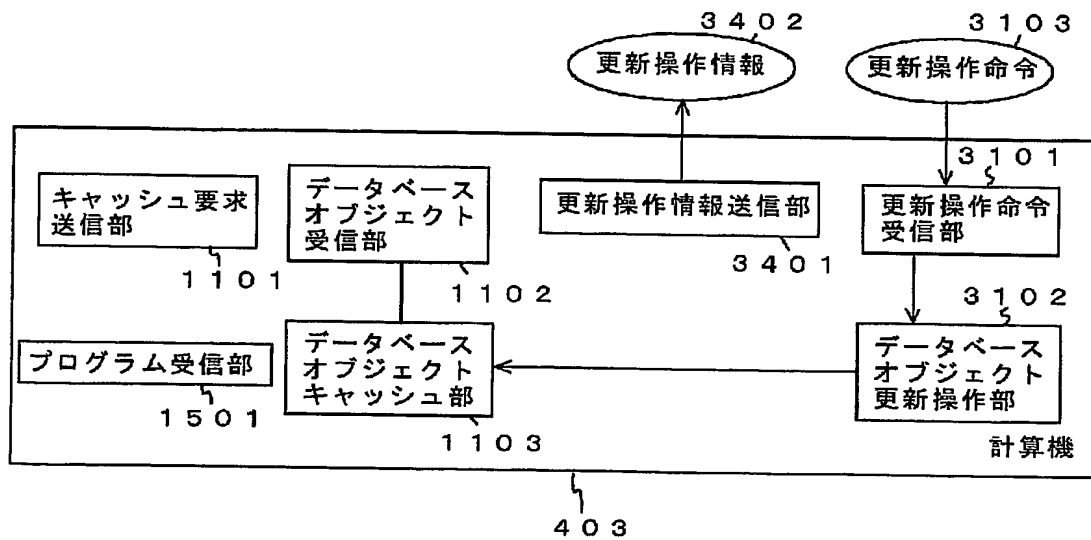
【図 3 1】



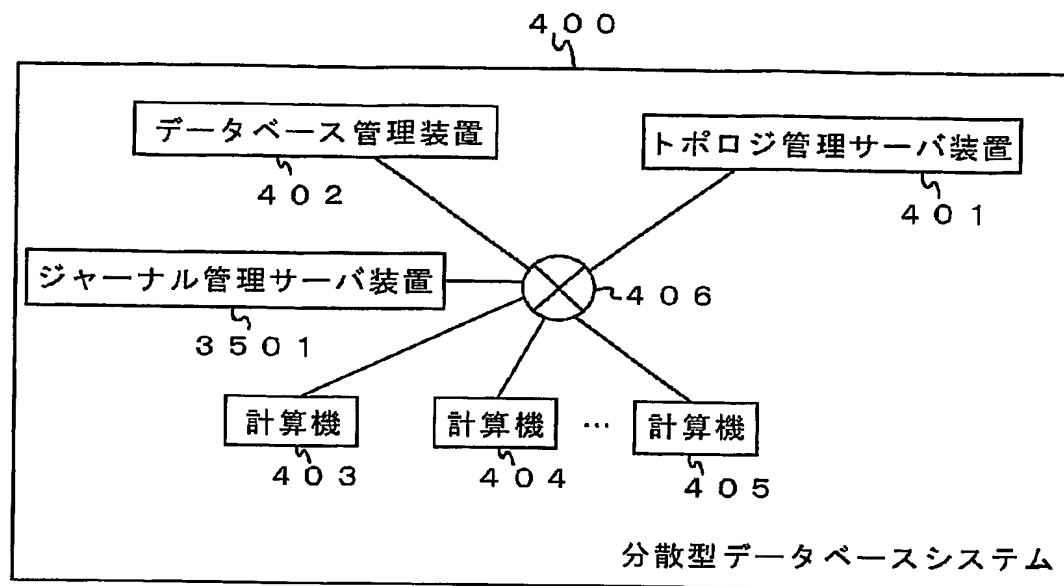
【図 33】



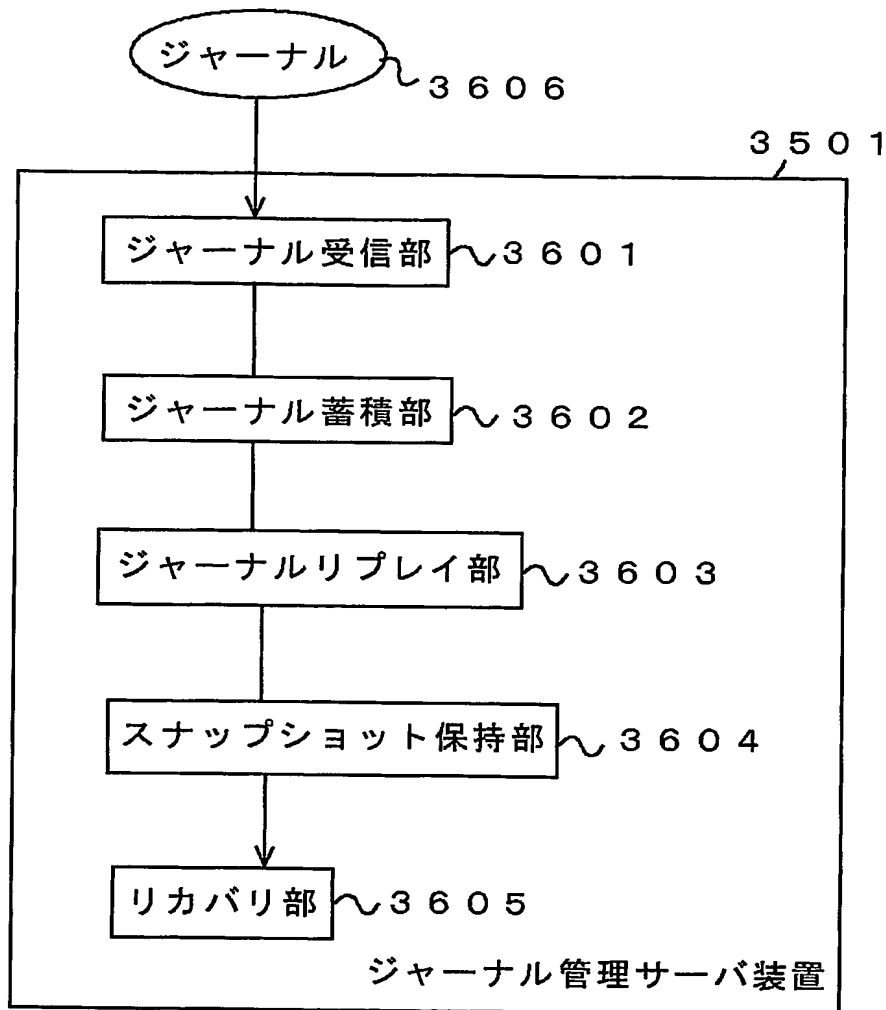
【図 34】



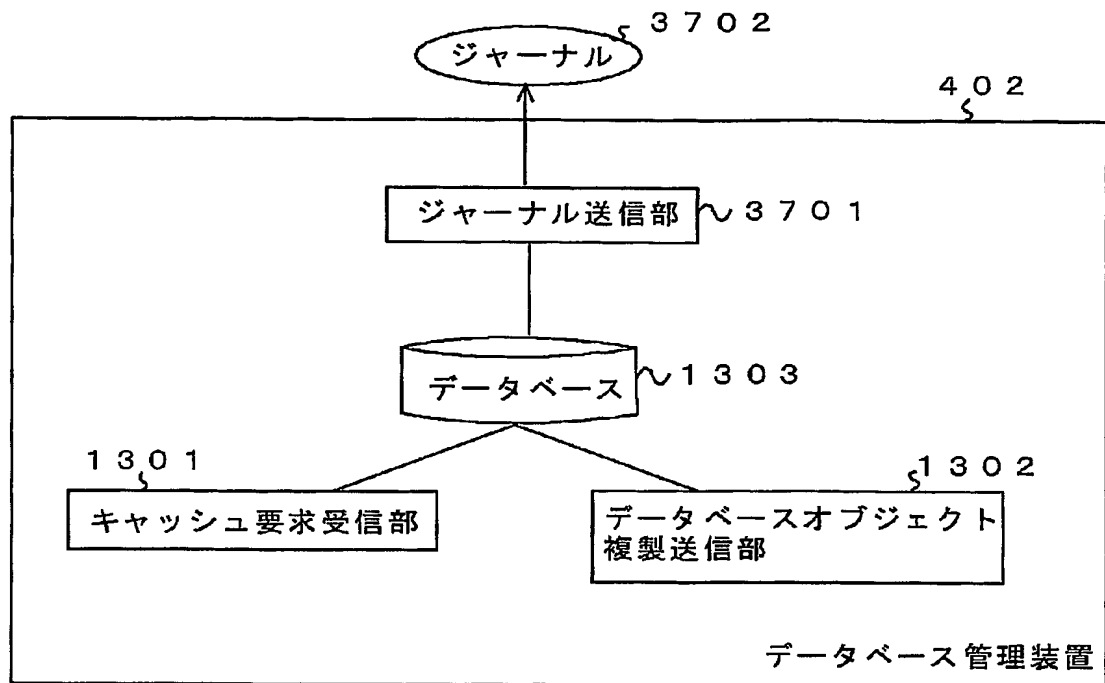
【図 35】



【図 36】



【図 37】



【書類名】 要約書

【課題】 複数の計算機の余剰資源を有効に利用する。

【解決手段】 トポロジ管理サーバ装置 103 にて、データベース管理装置で管理されるデータベースオブジェクトを識別する識別子とそのデータベース管理装置を識別する識別子を対応付けて管理する。複数の計算機 104、105、…、106 がデータベースオブジェクトをキャッシュするためにデータベースオブジェクトを識別する識別子を含むキャッシュ要求をトポロジ管理サーバ装置 103 へ送信すると、トポロジ管理サーバ装置 103 は、データベース管理装置へキャッシュ要求を転送し、データベースオブジェクトが計算機へ送信され、そこでキャッシュされるようにする。計算機で、キャッシュされたデータベースオブジェクトを参照しながら計算を行なうプログラムを動かしてもよい。

【選択図】 図 1

認定・付加情報

特許出願の番号	特願 2 0 0 3 - 0 1 2 5 4 5
受付番号	5 0 3 0 0 0 9 0 3 3 8
書類名	特許願
担当官	第七担当上席 0 0 9 6
作成日	平成 1 5 年 1 月 2 4 日

< 認定情報・付加情報 >

【提出日】 平成15年 1月21日

次頁無

【書類名】 出願人名義変更届
【あて先】 特許庁長官殿
【事件の表示】
【出願番号】 特願2003- 12545
【承継人】
【持分】 50/100
【住所又は居所】 東京都港区芝四丁目 1 1 番 1 号
【氏名又は名称】 株式会社クエスト
【代表者】 野中 攻
【承継人代理人】
【識別番号】 100109553
【弁理士】
【氏名又は名称】 工藤 一郎
【手数料の表示】
【予納台帳番号】 100322
【納付金額】 4,200円
【提出物件の目録】
【物件名】 承継人であることを証する書面 1
【援用の表示】 追って手続補足書で提出する
【物件名】 代理権を証明する書面（委任状） 1
【援用の表示】 追って手続補足書で提出する

認定・付加情報

特許出願の番号	特願 2003-012545
受付番号	50301736306
書類名	出願人名義変更届
担当官	小野寺 光子 1721
作成日	平成16年12月27日

< 認定情報・付加情報 >

【提出日】 平成15年10月20日

【承継人】

【識別番号】 502152883

【住所又は居所】 東京都港区芝浦一丁目12番3号

【氏名又は名称】 株式会社クエスト

【承継人代理人】 申請人

【識別番号】 100109553

【住所又は居所】 東京都千代田区有楽町1丁目7番1号 有楽町電
気ビル南館9階

【氏名又は名称】 工藤 一郎

特願 2003-012545

出願人履歴情報

識別番号

[503030067]

1. 変更年月日

2003年 1月21日

[変更理由]

新規登録

住所

千葉県浦安市明海14番地海園の街2-805号

氏名

西山 修平

出 願 人 履 歴 情 報

識別番号

[502152883]

1. 変更年月日

2002年 4月12日

[変更理由]

新規登録

住 所

東京都港区芝四丁目11番1号 CBC田町ビル

氏 名

株式会社クエスト

2. 変更年月日

2003年11月12日

[変更理由]

住所変更

住 所

東京都港区芝浦一丁目12番3号

氏 名

株式会社クエスト

Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/JP03/014390

International filing date: 12 November 2003 (12.11.2003)

Document type: Certified copy of priority document

Document details: Country/Office: JP
Number: 2003-12545
Filing date: 21 January 2003 (21.01.2003)

Date of receipt at the International Bureau: 27 January 2005 (27.01.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)



World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse